Niveau(x) d'enseignement : 0 Numéro EDUQ : 9270280
Centre d'intérêt : 540

Scallon, Gérard.

La construction d'un test diagnostique selon des facettes. 1, Présentation d'un modèle de recherche pédagogique / Gérard Scallon. -- [Québec] : Université Laval, Département de mesure et évaluation, c1981.
ii, 65 p. : ill. ; 28 cm. -- (Monographies en mesure et évaluation)

Code: RPDP-81-218. Bibliogr. : p. 61-63.

Vedette(s)-secondaire(s): Université Laval. Département de mesure et évaluation

---

Descripteur(s): Elaboration de tests * Elaboration d'examens * Test de diagnostic * Instrument d'évaluation * Evaluation formative

Identificateur(s) de sujet: Evaluation diagnostique

Le modèle de recherche pédagogique suggéré ici s'inscrit dans le domaine du diagnostic pédagogique et implique la construction systématique d'une épreuve de rendement selon un schème à deux ou plusieurs facettes dans le but de fractionner en plusieurs profils les scores observés dans un groupe d'élèves. Les principaux éléments du modèle sont présentés: la théorie des facettes et la notion de profil diagnostique, l'étude de variations introduites dans des questions d'examen et l'analyse des composantes de la variance associée à un schème à facettes multiples.

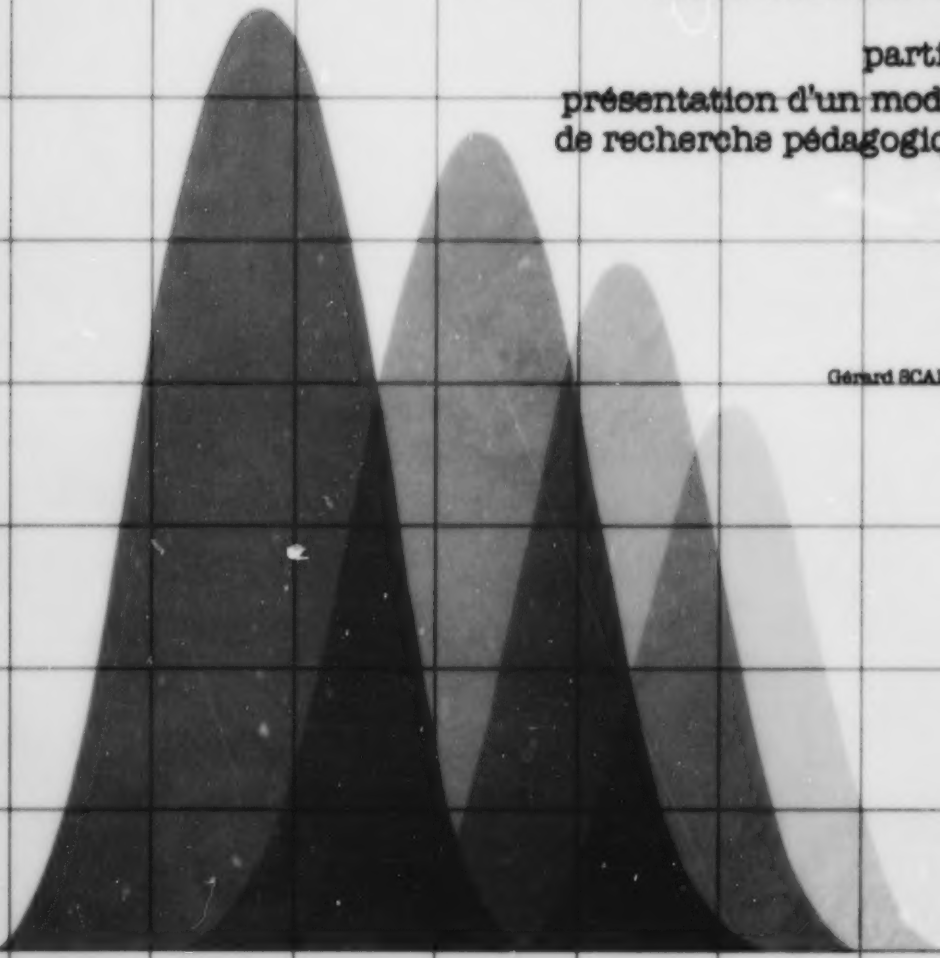
Type de document :

MONOGRAPHIES EN MESURE ET ÉVALUATION

LA CONSTRUCTION D'UN TEST DIAGNOSTIQUE SELON DES FACETTES

partie I

présentation d'un modèle de recherche pédagogique



Gérard SCALLON

DÉPARTEMENT DE MESURE ET ÉVALUATION
FACULTÉ DES SCIENCES DE L'ÉDUCATION

UNIVERSITÉ LAVAL

9270280

LA CONSTRUCTION D'UN TEST DIAGNOSTIQUE SELON DES FACETTES

<u>partie I</u>: présentation d'un modèle de recherche pédagogique

Gérard SCALLON

(professeur au Département de mesure et évaluation, Faculté des sciences de l'éducation, Université Laval)

© DEPARTEMENT DE MESURE ET EVALUATION
UNIVERSITE LAVAL, octobre 1981

LA CONSTRUCTION D'UN TEST DIAGNOSTIQUE SELON DES FACETTES

TABLE DES MATIERES

Partie I: présentation d'un modèle de recherche pédagogique

page

INTRODUCTION 1

Chapitre I: Les instruments de l'évaluation formative dans le cadre du diagnostic pédagogique 5

Chapitre II: La construction systématique d'une épreuve de rendement selon la théorie des facettes 13

Chapitre III: Recherches traitant de variations introduites dans des questions d'examen 27

Chapitre IV: L'analyse des composantes de la variance dans une perspective de diagnostic pédagogique 33

CONCLUSION A LA PREMIERE PARTIE 57

BIBLIOGRAPHIE (première partie) 61

LISTE DES TABLEAUX

1- Notes obtenues par 10 élèves pour une composition littéraire évaluée par 5 enseignants 35

2- Analyse de la variance des notes attribuées par cinq enseignants (juges) à dix compositions littéraires (élèves) 36

Liste des tableaux (suite)

3- Estimés des composantes de la variance et de leurs contributions relatives en pourcents en rapport avec l'exemple fictif des cinq juges qui ont évalué dix compositions littéraires 37

4- Succès (ou échecs) obtenus par 10 étudiants fictifs à une épreuve composés de 9 questions rédigées selon deux facettes croisées A et B 41

5- Profils des élèves en ragard de chacune des facettes à partir des résultats apparaissant au tableau 4 42

6- Estimés des composantes de la variance en rapport avec les résultats fictifs à une épreuve de 9 questions rédigées selon deux facettes croisées A et B 45

7- Présentation de quelques patrons possibles de contributions relatives des composantes de la variance reliées à une facette A, dans un schème S(sujets) x A x B avec une observation par cellule, et interprétation du point de vue diagnostique 54

LISTE DES FIGURES

1- Illustration d'un test centré sur un objectif pour mesurer la capacité d'additionner deux nombres présentés horizontalement 8

2- Echantillon de dessins variant d'après la grandeur, la forme, le signe intérieur et le contour 17

3- Schème à quatre facettes (ou dimensions) complètement croisées pour représenter les variations entre les dessins de la figure précédente 18

4- Profils pouvant être obtenus en groupant les seize dessins en divers sous-tests selon chacune des facettes introduites 22

APPENDICE: Carrés moyens attendus et composantes de la variance estimées à partir des carrés moyens observés pour le schème p x a x b dans lequel p (sujets) est aléatoire simple, a et b sont deux facettes du test qui correspondent à des dimensions fixées 65

INTRODUCTION

Le modèle de recherche pédagogique qui est suggéré dans cet ouvrage s'inscrit dans le domaine du diagnostic pédagogique. Ce modèle implique la construction systématique d'une épreuve de rendement selon un schème à deux ou plusieurs facettes dans le but de fractionner en plusieurs profils les scores observés dans un groupe d'élèves. La construction d'une épreuve selon des facettes s'applique à une épreuve de rendement centrée sur un seul objectif et présente ainsi une façon particulière de concevoir un test de maîtrise.

Dans le cas d'un objectif pédagogique reconnu comme étant difficile à atteindre par des élèves, les profils qui se dégagent d'un schème à facettes multiples devraient idéalement décrire les symptômes de la non maîtrise de cet objectif dans un groupe d'élèves. Pour qu'une épreuve construite selon des facettes ait une portée diagnostique il faut établir le degré de différenciation en difficulté de chacune des facettes introduites au moyen de procédés statistiques appropriés. L'analyse des composantes de la variance associées aux élèves et aux facettes de l'épreuve semble être un procédé tout indiqué pour ce genre d'étude puisqu'elle apporte des éléments de réponse aux questions suivantes en regard de l'une ou l'autre des facettes introduites:

a) cette facette est-elle associée à une variation importante dans la difficulté des questions qui composent l'épreuve?

b) la variation en difficulté, introduite par cette facette, est-elle conditionnée ou non par l'introduction d'autres facettes?

c) la variation en difficulté, introduite par cette facette, est-elle généralisable à tous les élèves d'une population donnée ou n'affecte-elle que certains élèves?

Les diverses réponses à ces questions feront que la facette choisie comme cible pourra être qualifiée ou non de catégorie diagnostique en ce sens que les profils d'élèves établis selon cette facette présenteront ou non des symptômes fiables de la non maîtrise d'un objectif pédagogique. Les conditions auxquelles l'influence d'une facette peut être soumise permettront de

nuancer l'interprétation des symptômes observés et de suggérer des stratégies de correctif pédagogique appropriées aux circonstances dans lesquelles la non maîtrise d'un objectif aura été constatée.

Le modèle de recherche pédagogique qui est présenté dans cet ouvrage est au carrefour de trois approches qui, tout en étant distinctes, n'en présentent pas moins des préoccupations communes. En premier lieu, la construction systématique d'une épreuve de rendement est une idée empruntée à la théorie des facettes de Guttman (1954). En second lieu, l'étude de l'influence de certaines variations introduites dans des problèmes ou dans des questions d'examen est la caractéristique de plusieurs recherches de nature psychopédagogique ou de nature didactique qui ont été publiées dans la littérature. En dernier lieu, l'analyse des composantes de la variance est un élément important d'un modèle d'analyse statistique appelé "théorie de la généralisabilité". Cette théorie, développée par Cronbach, Gleser, Nanda et Rajaratnam (1972), est une extension de la conception classique de la fiabilité appliquée au domaine de la mesure.

Le modèle de recherche pédagogique proposé ici est une combinaison de ces trois approches dans le but d'analyser les circonstances particulières dans lesquelles la non maîtrise d'un objectif pédagogique peut être observée. Les recherches de nature psychopédagogique ou de nature didactique fournissent au modèle un contenu et des hypothèses à vérifier quant aux facettes à introduire. La théorie des facettes vient encadrer le domaine d'intérêt et définit un schème expérimental pour la rédaction des questions, schème qui s'apparente aux schèmes factoriels utilisés en analyse classique de la variance. La théorie de la généralisabilité, qui a pour fondements l'analyse de la variance, présente des éléments rationnels qui sont compatibles avec l'analyse statistique de questions d'examen rédigées selon des facettes. Dans un souci de simplification seuls les estimés des composantes de la variance et leurs contributions relatives en pourcents ont été retenus comme moyens d'analyse d'une épreuve de rendement construite selon des facettes. Ceci n'exclut pas que des raffinements puissent être ajoutés avec l'introduction de coefficients de généralisabilité. Ces raffinements seront d'ailleurs signalés à certaines occasions.

Le présent ouvrage s'adresse à toute personne intéressée au diagnostic pédagogique. [La construction d'une épreuve de rendement selon des facettes n'exige pas de connaissances particulières si ce n'est qu'une bonne dose de pédagogie, une certaine expérience de l'enseignement et un souci d'innovation pédagogique.] On ne peut en dire autant cependant de l'analyse statistique proposée pour compléter ce mode de construction d'une épreuve de rendement. La théorie de la généralisabilité exige une bonne formation en analyse de la variance. Pour maintenir l'intérêt du lecteur non initié à ce genre d'analyse, les aspects les plus techniques ont été réduits à leur plus simple expression sans pour autant négliger les éléments les plus utiles à une bonne compréhension, éléments qui seront signalés au besoin. Il sera d'ailleurs démontré avec quelques exemples qu'il est possible de prévoir certaines conclusions avec un traitement statistique des plus élémentaires: ce sera en quelque sorte l'examen "à l'oeil nu" des résultats à une épreuve de rendement, examen guidé par une certaine intuition. C'est alors que l'on pourra apprécier l'analyse des composantes de la variance car des situations concrètes feront voir que cette technique statistique permet de décrire des symptômes qui ne sont pas toujours visibles à l'oeil nu.

Pour des raisons de commodité, l'ouvrage a été divisé en deux parties. La première partie est consacrée à la présentation du modèle avec ses principaux éléments: la théorie des facettes et la notion de profil diagnostique, l'étude de variations introduites dans des questions d'examen et l'analyse des composantes de la variance associée à un schème à facettes multiples. La seconde partie traitera des applications du modèle proposé au moyen d'instruments de mesure déjà rédigés selon des facettes et pour lesquels des résultats d'analyse de la variance étaient déjà accessibles au moment de rédiger cet ouvrage. Dans cette seconde partie de l'ouvrage, il sera question également de divers aspects d'ordre méthodologique reliés à l'analyse statistique proposée pour interpréter les résultats à un test construit selon des facettes. Rappelons que le présent texte ne traitera que de la partie I de l'ouvrage.

CHAPITRE I

LES INSTRUMENTS DE L'EVALUATION FORMATIVE DANS LE CADRE DU DIAGNOSTIC PEDAGOGIQUE

Les instruments de mesure utilisés pour réaliser l'évaluation formative de l'apprentissage ont été désignés par des appellations variées: tests centrés sur un critère (criterion-referenced tests), tests centrés sur un ou plusieurs objectifs (objectives-based tests) ou tests de maîtrise. La réalité que ces diverses appellations sous-tendent a cependant subi une certaine évolution depuis l'avènement de la mesure dite "critériée". Les impératifs du diagnostic pédagogique ne sont pas étrangers à cette évolution. Dans ce chapitre nous tenterons de décrire les principaux raffinements qui ont marqué l'évolution du concept de mesure critériée dans le cadre de l'évaluation formative en général ainsi que les besoins en instrumentation imposés par le diagnostic pédagogique. Nous terminerons ce chapitre en précisant le contexte, la nature et les limites du diagnostic pédagogique tel qu'entendu dans cet ouvrage.

<u>Les tests de maîtrise</u>.-

La mesure et l'évaluation de l'apprentissage, qui ont été modelées pendant plusieurs années sur la psychométrie, ont connu des développements majeurs depuis l'avènement de la pédagogie de la maîtrise. Presque partout où l'accent a été mis sur la poursuite et l'atteinte d'objectifs pédagogiques clairement définis, la mesure dite critériée est apparue comme un type de mesure plus approprié que la mesure normative. Dans les premiers textes publiés sur le sujet, la mesure critériée se définissait principalement par le caractère non normatif de l'interprétation des résultats (voir Popham et Husek, 1969 et Glaser et Nitko, 1971). Toutefois, le besoin de recourir à un cadre de référence bien défini pour interpréter des résultats d'examen, en remplacement d'une performance de groupe, et qui s'ajoute au caractère non normatif de l'interprétation des résultats, a été souligné à plusieurs reprises.

Ebel (1962, p. 15) a utilisé l'expression "content-standard" pour désigner des tests permettant d'indiquer le pourcentage de tâches réussies dans un échantillon provenant d'un univers de tâches clairement défini. Airasian et Madaus (1972) ont particulièrement insisté sur le caractère incomplet de l'interprétation non normative des résultats pour définir la mesure dite critériée (pp. 3 et 4); pour Airasian et Madaus, les instruments de mesure en relation avec la mesure critériée doivent être axés sur une définition claire des objectifs pédagogiques. C'était d'ailleurs l'époque où sont apparus les tests centrés sur un ou plusieurs objectifs (objectives-based tests).

La pédagogie de la maîtrise n'a pas influencé uniquement la mesure proprement dite de l'apprentissage mais a également amené l'implantation d'un nouveau type d'évaluation pédagogique qu'on appelle "évaluation formative" (Bloom, Hastings et Madaus, 1971; Airasian et Madaus, 1972; Bégin et Dussault, 1980).

L'évaluation formative se caractérise par un processus continu d'évaluation de l'apprentissage dont le rôle essentiel est de suivre et d'assurer le progrès continu de chaque élève. Ce rôle est assumé auprès de ce dernier en intervenant périodiquement dans son cheminement. Le constat de la non maîtrise d'un objectif chez un élève doit alors déclencher une intervention pédagogique de nature corrective, ce qui a pour conséquence d'interrompre momentanément la progression de l'élève vers de nouveaux objectifs. On voit donc aisément que la pédagogie de la maîtrise et l'évaluation formative se fondent dans un même esprit. Les instruments de l'évaluation formative doivent donc non seulement se prêter à une interprétation non normative des résultats, doivent non seulement être conçus en regard d'un objectif cible clairement explicité mais doivent aussi posséder un caractère diagnostique. C'est sans doute pour répondre à cette triple exigence de l'évaluation formative que Hively (1974) a introduit la notion de mesure centrée sur un domaine. Le caractère incomplet du test centré sur un objectif, aux fins du diagnostic pédagogique, a d'ailleurs été souligné à quelques reprises dans la littérature reliée à l'évaluation formative (par exemple, Baker, 1974, p. 11; Millman, 1974, pp. 324-325).

<u>La faible valeur diagnostique du test centré sur un objectif: un exemple</u>.-

Supposons que l'on ait un objectif pédagogique qui précise que l'élève devra être capable d'additionner deux nombres présentés horizontalement. Cet objectif s'inscrit dans le vaste domaine des problèmes d'addition mais ses spécifications nous limitent à des problèmes d'addition de <u>deux nombres</u> présentés <u>horizontalement</u>. Une épreuve de rendement conçue pour mesurer le degré d'atteinte de cet objectif pourrait ressembler à celle qui est reproduite à la figure 1. On remarquera que chacune des questions (ou chacun des problèmes) est congruente à l'objectif explicité (chacune présente deux nombres disposés horizontalement). Pourtant, en dépit de la clarté de l'objectif, le test donné en exemple présente une foule d'ambiguïtés du point de vue diagnostique. En présence d'un élève qui ne réussirait que quatre des dix problèmes du test, par exemple, que peut-on affirmer au sujet des difficultés qu'il éprouve? S'agit-il d'une déficience généralisée à l'égard de tout problème d'addition? S'agit-il de difficultés particulières que cet élève éprouve devant certains problèmes d'addition et non devant d'autres problèmes?

Un examen de chacun des dix problèmes qui composent l'épreuve donnée en exemple à la figure 1 permettra de mettre en évidence certaines variations entre ces problèmes. Par exemple, il y a des problèmes qui n'impliquent aucune retenue (les numéros 1, 2, 3, 5 et 10), les autres problèmes impliquant au moins une retenue. Parmi les problèmes impliquant au moins une retenue, les problèmes 6 et 9 impliquent deux retenues en succession et le problème 6, par surcroît, exige la manipulation d'un zéro dans le cas d'une retenue. De plus, le repérage et l'alignement des unités et des dizaines (ou des centaines s'il y a lieu) peuvent présenter une difficulté additionnelle à des élèves déficients, dans les problèmes où la quantité de chiffres n'est pas la même dans les deux nombres à additionner (les numéros 2, 3, 4 et 10). Parmi toutes les variations que l'on pourrait détecter entre les problèmes il serait difficile de décrire celle ou celles qui présentent une difficulté accrue pour un élève qui ne maîtrise pas l'objectif mesuré par le test. La réponse donnée à chaque problème indiquerait les problèmes réussis et les problè-

Effectue les additions suivantes et écris ta réponse dans l'espace qui est à droite de chaque problème.

(1)	31 + 13	______
(2)	51 + 4	______
(3)	35 + 3	______
(4)	43 + 8	______
(5)	24 + 11	______
(6)	294 + 106	______
(7)	456 + 128	______
(8)	325 + 152	______
(9)	437 + 298	______
(10)	450 + 40	______

Figure 1: illustration d'un test centré sur un objectif pour mesurer la capacité d'additionner deux nombres présentés horizontalement.

mes échoués par un élève qui aurait obtenu une note faible au total mais les conclusions qu'on pourrait en tirer seraient basées sur un nombre relativement faible d'occurences. Par exemple l'échec du problème 4 (qui demande la somme de 43 et 8) traduit-il une difficulté au niveau de l'alignement des chiffres, une difficulté au niveau de la retenue ou une difficulté au niveau de connaissances de base proprement dites (savoir que la somme de 8 et 3 est 11)? Sur le plan de la fiabilité, l'échec du problème 6 est-il suffisant à lui-seul pour inférer, de façon sûre, une difficulté dans la manipulation d'un zéro dans le cas d'une retenue? De pareilles questions se posent à toute fin pratique à l'égard de chacun des problèmes qui serait échoué par un élève.

On constate aisément que, bien que composé de questions congruentes à un objectif clairement explicité, le test donné en exemple ne permet pas de circonscrire avec exactitude et avec une certaine fiabilité la ou les difficultés éprouvées par un élève qui ne maîtrise pas cet objectif. L'administration clinique du test, c'est-à-dire, l'observation directe des procédés de solution employés par certains élèves, révèlerait sans aucun doute des informations utiles au diagnostic pédagogique. On pourrait y arriver également au moyen d'une codification appropriée des réponses données à certains problèmes. [Cependant, la multiplicité des interventions à réaliser dans un processus d'évaluation formative rendent très onéreux le recours à l'observation directe ou l'emploi de procédés analytiques de correction. Il faut donc se tourner du côté de la construction de tests pouvant être administrés collectivement tout en se prêtant à des procédés de correction abordables au point de pouvoir être utilisés fréquemment dans une classe.] Le problème ou le défi à relever est de développer des instruments de mesure qui, tout en respectant ces contraintes pratiques, livreront toutes les informations utiles à la conduite du diagnostic pédagogique en milieu scolaire.

Nature et limites du diagnostic pédagogique.-

L'exemple que nous venons de donner nous amène à nous interroger sur la nature du diagnostic pédagogique d'après le sens qui lui sera donné tout au long de ce document. La question est de savoir en quoi l'étude de varia-

tions à introduire dans des questions d'examen ou dans des problèmes à résoudre peut intéresser le diagnostic pédagogique.

Le diagnostic se définit par l'action de déterminer une maladie d'après ses symptômes. Cette définition, qui est donnée dans le petit Robert, rejoint la conception populaire de ce qu'est le diagnostic en général. Devant une forte fièvre (symptôme) on soupçonnera une affection virale (explication causale). En présence de plusieurs ratés successifs d'un moteur (symptôme) on vérifiera si une panne d'essence en est l'origine (explication causale). En observant certaines indications anormales d'un extrant d'ordinateur (symptôme) on cherchera du côté de certaines erreurs qui ont pu se glisser au moment de la programmation (explication causale). Il semble donc que le diagnostic en général soit un processus qui consiste à observer et à noter des symptômes et à en dégager une ou des explications causales.

Dans le cas du rendement scolaire et plus particulièrement lorsqu'il est question de vérifier chez un élève la maîtrise d'un objectif, les explications causales d'une déficience éventuelle ne se présentent pas aussi naturellement que dans le cas des quelques exemples énumérés précédemment. Une déficience en rendement scolaire peut provenir d'une multitude de causes. Certaines de ces causes sont inhérentes au processus même d'enseignement et d'apprentissage (temps d'apprentissage trop court, exercices en nombre insuffisant, explications non appropriées, mise en situation inadéquate etc.) alors que d'autres causes, plus lointaines, sont du domaine de la préparation antérieure, des aptitudes, de la motivation ou des influences multiples de l'environnement passé et présent de l'élève. Anna Bonboir (1970(a), p. 60) a proposé une classification des manifestations possibles de l'échec selon que les déficiences observées chez un élève sont généralisées à plusieurs apprentissages ou sont spécifiques à un secteur donné ou encore selon que les déficiences observées sont communes à tout un groupe d'élèves ou sont caractéristiques de quelques élèves seulement. Scallon (1980, p. 7) a repris substantiellement le même schéma de classification des diverses éventualités pouvant entourer l'échec par un élève dans l'atteinte d'un objectif pédagogique et ce, dans le but d'expliciter un ensemble de prises de décision parmi lesquelles se situe le correctif pédagogique. Il faut comprendre que l'observation d'un échec dans l'atteinte d'un objectif, comme événement isolé, ne

conduit pas à un diagnostic précis puisque d'après le schème proposé par Anna Bonboir il faut connaître l'histoire individuelle des apprentissages de l'élève et la fréquence de maîtrise de l'objectif dans un groupe donné. De plus, la classification proposée en est une de symptômes. Les explications causales des diverses éventualités sont de l'ordre des inférences. Par exemple si la non maîtrise d'un objectif est la caractéristique de tout un groupe d'élèves, on remettra alors en question la pédagogie. Si par contre la non maîtrise de l'objectif est la caractéristique de quelques élèves seulement et que ces élèves éprouvent des difficultés depuis longtemps, on cherchera alors du côté des aptitudes ou du côté de la motivation de ces élèves à l'égard de la matière visée par cet objectif. Enfin, si la non maîtrise de l'objectif n'est la caractéristique que de quelques élèves seulement et que la difficulté n'est que passagère pour ces élèves on invoquera alors des causes inhérentes à la situation même d'enseignement et d'apprentissage: c'est d'ailleurs à ce niveau que se situe le correctif pédagogique proprement dit et qui pourrait se traduire par un traitement individualisé.

Qu'en est-il de l'étude des variations pouvant être introduites entre les questions ou problèmes qui composent une épreuve de rendement destinée à vérifier l'atteinte d'un objectif pédagogique par un élève ou par un groupe d'élèves? Cette approche à elle-seule ne saurait être qualifiée de diagnostique. Elle se limite tout au plus à circonscrire les circonstances particulières (types de questions ou de problèmes) dans lesquelles un ou plusieurs élèves peuvent afficher leur non maîtrise de l'objectif. L'approche se présente donc comme une manière de décrire les symptômes accompagnant la non maîtrise d'un objectif. En se rappelant l'exemple qui a été donné précédemment avec des problèmes d'addition de deux nombres disposés horizontalement, le fait d'observer qu'un élève réussit moins bien les problèmes avec retenue que les problèmes sans retenue est une façon de circonscrire la non maîtrise de l'objectif de la part de cet élève. On comprendra toutefois que cette observation n'est pas à proprement parler un diagnostic car plusieurs causes peuvent être invoquées au départ pour expliquer une telle éventualité. Ce n'est qu'à la lumière de l'histoire personnelle de l'élève et de la fréquence d'occurence du même phénomène dans un groupe d'élèves que nous pourrons risquer certaines hypothèses qui seront de nature diagnostique.

L'étude des variations entre les questions d'un examen dans le but de circonscrire la non maîtrise éventuelle d'un objectif n'a pas fait l'objet d'une insistance particulière dans la littérature en rapport avec le diagnostic pédagogique. Millman (1974), Hively (1974) et Baker (1974) y ont fait brièvement allusion et guère plus. Pourtant, de nombreuses études de nature didactique, notamment dans le domaine des mathématiques, ont démontré l'impact que pouvait avoir notre connaissance que certains types de problèmes sont plus difficiles que d'autres sur l'analyse psychopédagogique d'une matière. Quelques unes de ces études seront citées dans un autre chapîttre de ce document. L'étude de l'influence de certaines variations pouvant être introduites dans des questions qui composent une épreuve de rendement, et ce au moyen de facettes, ne constitue donc qu'un seul élément d'un processus complexe qui est le diagnostic pédagogique.

CHAPITRE II

LA CONSTRUCTION SYSTEMATIQUE D'UNE EPREUVE DE RENDEMENT SELON LA THEORIE DES FACETTES

Le test centré sur un objectif, c'est-à-dire un test composé de plusieurs questions congruentes à un même objectif, présente des lacunes importantes lorsqu'il s'agit de décrire les circonstances particulières dans lesquelles cet objectif n'est pas maîtrisé par un élève ou par un groupe d'élèves. Cette affirmation a quelque peu été étayée au moyen d'un exemple impliquant des additions de nombres disposés horizontalement. Dans cette perspective, la mesure centrée sur un domaine est apparue comme étant une approche susceptible d'être orientée vers le diagnostic pédagogique beaucoup plus que le test centré sur un ou plusieurs objectifs. Les quelques recherches menées dans cette direction (Hively, Patterson et Page, 1968; Macready et Merwin, 1973) indiquent qu'il est possible de fractionner un domaine en plusieurs sous-domaines donnant lieu chacun à un test composé de questions très homogènes entre elles. Les divers sous-scores obtenus par chaque élève établissent un profil de compétence à l'égard d'une habileté mesurée. La pertinence du profil, aux fins du diagnostic pédagogique, peut être évaluée au moyen d'une recherche dans laquelle l'homogénéité des divers sous-tests est établie comme dans l'étude de Macready et Merwin (1973) ou au moyen de l'analyse des composantes de la variance associées aux questions d'un même sous-test ou associées aux sous-tests eux-mêmes comme dans l'étude de Hively, Patterson et Page (1968).

Pour ce qui est d'un test de rendement conçu à des fins de diagnostic pédagogique, le modèle de construction et d'analyse qui est présenté dans cet ouvrage s'inspire en grande partie des recherches de Macready et Merwin et de Hively, Patterson et Page mais propose une procédure plus systématique de définition d'un domaine, procédure qui a été empruntée à la théorie des facettes de Guttman. Les applications possibles de cette théorie à la construction d'une épreuve de rendement de même que la notion de profil diagnostique qui s'en dégage seront présentées dans ce chapitre.

<u>La théorie des facettes</u>.-

L'idée que des variables puissent être manipulées et combinées dans un schème d'analyse n'est guère nouvelle. On la retrouve au départ dans les schèmes d'analyse de la variance avec plan factoriel dans lesquels deux ou plusieurs dimensions sont introduites pour catégoriser des observations. Le sexe, la catégorie d'âge, le niveau socio-économique et le traitement dispensé sont des exemples de telles dimensions. Chaque dimension peut avoir deux ou plusieurs niveaux. Par exemple, "masculin" et "féminin" sont les deux niveaux de la dimension "sexe"; de même, "jeune", "d'âge moyen" et "âgé" pourraient être les trois niveaux de la dimension "catégorie d'âge". Ainsi, chacune des observations recueillies dans un plan factoriel peut être décrite comme étant caractérisée par un niveau de chaque dimension.

Guttman (1954) a proposé un modèle de construction de tests qui est analogue aux schèmes d'analyse de la variance avec plan factoriel. Dans ce modèle, les dimensions sont appelées "facettes" et les niveaux des dimensions, "éléments". Ce modèle est appelé "facet design" et permet de définir un ensemble d'items ou de questions et d'en arriver à un tableau de spécifications à deux ou plusieurs dimensions. Chaque item d'un test construit selon cette approche peut être situé dans ce tableau et ainsi être caractérisé au moyen d'un élément de chaque facette.

Guttman lui-même n'a pas présenté de façon systématique sa théorie des facettes et les principes qui y sont rattachés sont disséminés dans de nombreuses publications. L'essentiel de cette approche à la construction des tests a été clairement décrit par Humphreys (1962) et par Millman (1974). De tous les texte consultés celui de Millman présente une définition qui est apparue comme étant la plus précise en ce qui a trait à la notion même de facette:

> Facets are dimensions of a situation believed relevant in the measurement effort. In the domain-referenced application, facets can be thought of as the dimensions or characteristics on which the items in the domain are to differ. Facets are chosen before an item domain is described, and, thus, they differ from factors that emerge after data are collected and analyzed, which occurs with factor analysis (1974, p. 332).

On voit donc, d'après cette définition, qu'une facette constitue une dimension selon laquelle des items ou questions d'un ensemble donné peuvent varier. Par exemple, un test d'addition peut contenir des problèmes qui varient selon le mode de présentation (disposition verticale ou disposition horizontale des nombres à additionner), selon la quantité de nombres à additionner (2, 3, 4 ou 5 par exemple) et selon qu'il n'y a aucune retenue à effectuer ou qu'il y a 1, 2, ou 3 retenues à effectuer. Les facettes qui auraient été introduites dans ce test d'addition seraient: (1) le mode de présentation, (2) la quantité de nombres à additionner et (3) le nombre de retenues à effectuer.

Les divers champs d'application de la théorie des facettes,ou plus précisément du schème à facettes multiples, à la construction d'instruments de mesure ont d'abord été celui de la mesure des attitudes et ensuite ceux de la mesure de l'intelligence et de la mesure du rendement (Guttman et Schlesinger, 1966 et 1967). A l'origine de cette théorie la préoccupation majeure était d'étudier la structure des intercorrélations entre les items d'un test construit selon des facettes. Bien que Guttman et Schlesinger aient suggéré des applications possibles de la théorie des facettes au diagnostic en général, la construction de tests de rendement selon un schème à facettes multiples est demeurée une entreprise marginale dans les écrits traitant de l'évaluation pédagogique. Dans un texte consacré à la mesure critériée et aux nombreux problèmes méthodologiques qui s'y rattachent, Millman (1974) traite de la théorie des facettes comme étant une approche possible à la définition d'un domaine ou d'un univers d'items. Toutefois, la théorie des facettes ne se limite pas à circonscrire un domaine. L'approche peut déboucher sur des orientations de recherche en diagnostic pédagogique comme en témoigne cette suggestion de Millman:

> We suggest that a domain be defined by those facets and elements that make a difference in how the examinee responds. Thus, for example, if the color the beam balance is painted does not affect how the student answers a weight comparison question, then color can be held constant or ignored. We know of no reason why facets of a domain need to be hierarchically arranged (1974, p. 338).

Pour Millman, la détection de facettes pouvant correspondre à des différences en rendement est un sujet de recherche en soi. L'analyse de la régression multiple ou l'analyse des composantes de la variance menée dans l'esprit d'une étude de généralisabilité seraient alors des outils statistiques appropriés (Millman, 1974, p. 338).

Les facettes d'un test et la notion de profil.-

On entend par "facette" toute variation introduite au niveau de la rédaction des questions devant composer une épreuve de rendement. Cette définition s'applique à tout instrument de mesure et même à une épreuve composée uniquement de questions congruentes à un seul objectif. Cette définition peut également s'appliquer à un test déjà construit mais dans un tel cas, l'analyse sera limitée par la présence ou l'absence d'un certain nombre d'éléments non contrôlés à l'origine. Pour faire saisir la notion de facette appliquée à la construction d'un test un exemple sera utilisé, exemple qui est traité dans les trois prochaines figures.

La figure 2 contient un échantillon de dessins qui présentent entre eux certaines variations: on voit des grandes et des petites figures géométriques (variation en grandeur); il y a des triangles et des carrés (variation quant à la forme); le contour (plein ou pointillé) varie également entre les dessins; enfin, toutes les figures ne contiennent pas le même signe à l'intérieur.

Supposons que ces dessins soient des stimuli à être utilisés dans un test pour mesurer une habileté quelconque et que l'on ne prenne qu'un stimulus par question. Si toutes les variations observées entre les stimuli nous intéressent, une construction systématique du test nous amènerait à concevoir un tableau de spécicifications comme celui qui est reproduit à la figure 3. Ce tableau de spécifications s'apparente à un schème à facettes multiples ("facet design" dans la terminologie de Guttman). On remarque que les quatre facettes retenues (ou variations d'intérêt entre les dessins) sont complètement croisées puisque les dessins choisis se prêtent facilement à une telle combinaison. Un dessin peut être grand ou petit quelle que soit sa forme et quels que soient son contour et le signe intérieur. Il en est de même pour chacune des variations considérées à tour de rôle.

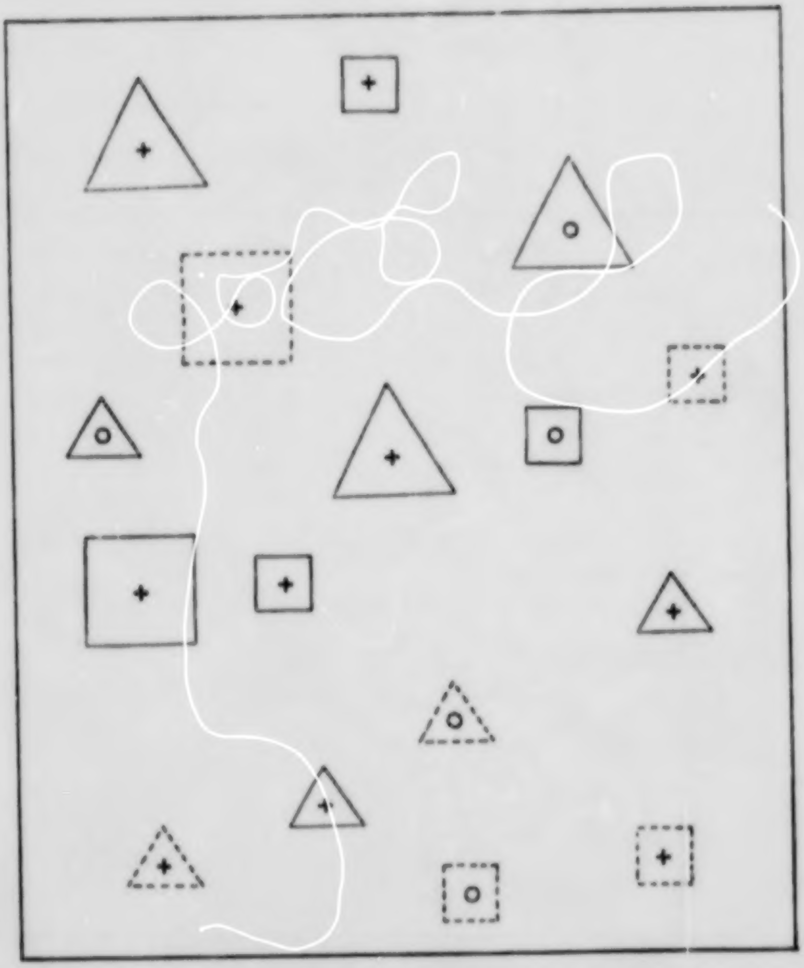


Figure 2 : échantillon de dessins variant d'après la grandeur, la forme, le signe intérieur et le contour.

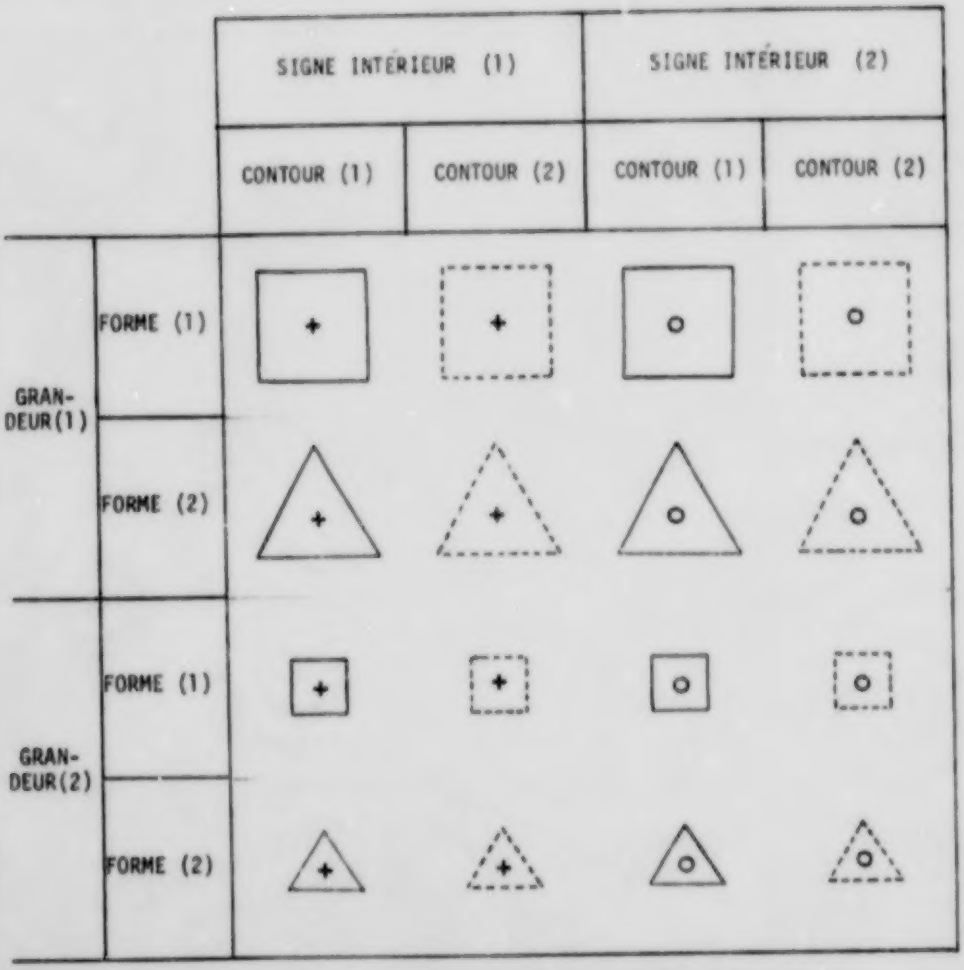


Figure 3 : schème à quatre facettes (ou dimensions) complètement croisées pour représenter les variations entre les dessins de la figure précédente.

On remarquera que certains dessins contenus dans le tableau de spécifications n'apparaissent pas dans l'échantillon de départ (figure 2). Ainsi, il n'y a aucun carré grand avec le signe intérieur "0" parmi les dessins de la figure 2. On remarquera également que plusieurs dessins apparaissent deux fois dans la figure 2 mais n'apparaissent qu'une seule fois dans le tableau de spécifications de la figure 3. Cette particularité tient au fait que le tableau de spécifications est le produit d'une construction systématique des dessins en vue de représenter le mieux possible les éléments d'un univers donné. L'exemple choisi illustre assez bien comment la théorie des facettes peut être utilisée à la définition d'un domaine.

Il convient de préciser immédiatement que le but visé par cette approche n'est pas de retenir les dessins qui présenteront un fort degré de discrimination entre les individus qui seront testés. La représentation fidèle de l'univers des dessins est primordiale ici, que certains dessins soient discriminatifs ou non.

La rédaction systématique des questions d'un examen selon un schème à facettes multiples se distingue des procédés connus jusqu'à ce jour pour fractionner un domaine. Selon ces procédés divers sous-tests auraient pu être envisagés avec l'exemple des dessins: les dessins de grandeur 1 dans un même sous-test, les dessins à contour plein dans un autre sous-test etc. Ce mode de fractionnement, peut-on le supposer, présente certaines lacunes:

a) le même dessin ou des dessins identiques risquent d'être répétés d'un sous-test à un autre et seraient ainsi éventuellement administrés à plusieurs reprises aux mêmes élèves;

b) parmi les dessins d'un sous-test, le risque est grand que certaines variations ne soient pas contrôlées;

c) il est difficile d'apprécier l'influence possible de la combinaison de certaines variations (interactions entre des facettes).

Le schème à facettes, utilisé dans la construction d'un test ou d'une épreuve de rendement, présente donc au départ une certaine rigueur qui d'ailleurs s'inspire des schèmes factoriels élaborés en analyse de la variance. Comme il sera démontré plus loin, le schème à facettes permet justement l'analyse des questions d'un examen au moyen de procédés statistiques qui sont du domaine de l'analyse de la variance. Sur le plan de l'utilisation des tests de maîtrise la théorie des facettes offre un double avantage pratique: la rigueur de l'échantillonnage des questions devant composer un examen fournit un appui à l'interprétation du score global obtenu par un élève en vue d'estimer son degré de maîtrise de l'habileté mesurée; en cas d'échec, les facettes déterminent des profils pouvant décrire les circonstances particulières dans lesquelles l'élève peut éprouver des difficultés. Les dernières lignes de ce chapitre seront consacrées à étayer ce deuxième point.

Si le tableau de spécifications obtenu avec les dessins, pour revenir à notre exemple, devait servir à élaborer un test pour établir la maîtrise ou la non maîtrise d'une habileté quelconque, il suffirait de préparer un certain nombre de questions pour chaque cellule du tableau (un même nombre de questions par cellule amènerait un test dit "balancé" ou équilibré); il faudrait ensuite déterminer un ordre d'apparition des questions dans le test et prévoir un seuil de performance. Le score global obtenu par un élève serait alors un estimé de sa compétence. Ce même score, comparé au seuil de performance, devrait permettre d'affirmer (non sans certains risques d'erreur) que l'élève maîtrise ou non l'habileté mesurée par le test.

Dans l'éventualité d'un échec on sera intéressé à décrire les circonstances particulières dans lesquelles l'élève semble éprouver des difficultés. C'est à ce niveau que le schème à facettes présente des avantages appréciables. Avec un tel schème il est possible de décomposer le score global de l'élève en profils, selon les dimensions mêmes du test c'est-à-dire selon les diverses variations ou facettes qui ont été introduites au moment de la rédaction des questions. Les interrogations qui se présentent d'elles-mêmes avec l'illustration des dessins sont les suivantes: un élève qui n'aurait que la moitié des points, par exemple, a-t-il mieux réussi les

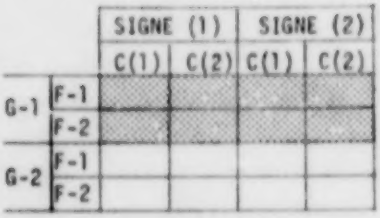
dessins de grandeur 1 que ceux de grandeur 2 (ou vice-versa)? A-t-il mieux réussi les dessins à contour plein que ceux à contour pointillé? Ou encore, d'une manière plus générale, quel est le type de dessin pour lequel l'élève a éprouvé le plus de difficulté? On comprendra que ce genre de question ne présente aucun intérêt si l'élève a échoué tous les problèmes ou presque ou si l'élève a réussi la majorité des problèmes du test. Les réponses à toutes ces questions peuvent être obtenues en examinant les divers profils décrivant la performance de l'élève aux divers types de problèmes. La figure 4 présente la correspondance entre le fractionnement du test en sous-tests de dessins selon le schème à facettes qui a été utilisé et ce, pour obtenir autant de profils qu'il y a de facettes. Si chaque case du tableau de spécifications apparaissant à la figure 3 correspond à une question du test, le fractionnement peut se faire en deux sous-tests d'égale grandeur pour chacune des facettes introduites. Dans l'exemple choisi, le fait d'avoir des facettes ayant deux éléments chacune et qui par surcroît sont complètement croisées a pour conséquence que les sous-tests obtenus sont tous de même grandeur (c'est-à-dire huit questions par sous-test si chaque case du tableau de spécifications correspond à une question). Cette situation particulière n'est cependant pas une exigence du modèle de construction et d'analyse d'une épreuve diagnostique tel que proposé dans ce document.

Avant de terminer ce chapître sur la construction d'un test d'après un schème à facettes multiples il convient de donner un exemple qui, cette fois, est beaucoup plus près de la réalité scolaire que ne l'est l'exemple des dessins. Dans une batterie de tests diagnostiques de français produits par l'A.I.E.S. (1980)[1] on trouve un test qui présente toutes les caractéristiques d'un instrument de mesure rédigé selon un schème à facettes. Il s'agit du test traitant de l'identification du sujet du verbe dans une phrase. Le test contient une vingtaine de phrases dans lesquelles le verbe est souligné. Pour chaque phrase l'élève doit écrire le ou les mots qui sont sujet du verbe qui est souligné. Les phrases qui composent ce test présentent plusieurs variations dont les deux suivantes:

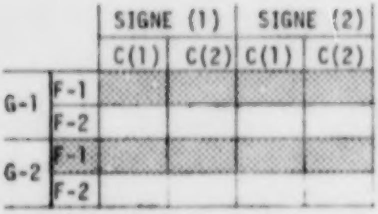
1) dans certaines phrases le sujet est placé avant le verbe et dans d'autres, le sujet est placé après le verbe;

1: Association des Institutions d'enseignement secondaire

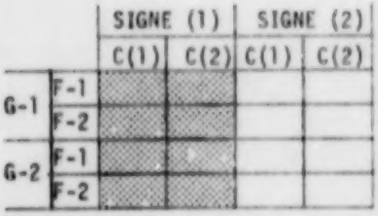
PROFILS



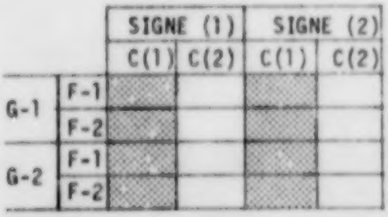
sous-test			prof.
GRANDEUR(1)			/8
GRANDEUR(2)			/8



sous-test			prof.
FORME(1)			/8
FORME(2)			/8



sous-test			prof.
SIGNE(1)			/8
SIGNE(2)			/8



sous-test			prof.
CONTOUR(1)			/8
CONTOUR(2)			/8

Figure 4 : profils pouvant être obtenus en groupant les seize dessins en divers sous-tests selon chacune des facettes introduites.

2) la nature des mots qui sont sujet du verbe souligné varie d'une phrase à l'autre: dans certaines phrases le sujet est un nom seul, pour d'autres phrases, le sujet est un pronom; le sujet peut également être représenté par plusieurs noms ou plusieurs pronoms ou par une combinaison de ces deux possibilités; enfin, dans quelques phrases, le sujet du verbe souligné est un verbe à l'infinitif.

Le test sur l'identification du sujet du verbe présente donc deux facettes: (1) la place occupée par le sujet par rapport au verbe souligné (avec ses deux éléments: avant ou après le verbe) et (2) la nature des mots qui sont sujet (avec six possibilités ou éléments). Ces deux facettes étant croisées on obtient douze situations-types représentées chacune par une case dans un tableau de spécifications à deux dimensions. Cependant, le cas de l'infinitif sujet placé après le verbe a été exclus ce qui réduit le nombre de situations-types à onze. La grille diagnostique qui accompagne ce test prévoit le fractionnement du score total de chaque élève en deux profils: le profil "position du sujet" avec deux sous-tests et le profil "mots qui sont sujet", avec six sous-tests. On trouvera une description plus détaillée de ce test dans Scallon (1980, pp. 22-25).

Les deux exemples qui viennent d'être donnés au sujet de la construction systématique d'un test selon un schème à plusieurs facettes ne manqueront pas de susciter plusieurs questions d'ordre théorique. La théorie des facettes peut devenir une approche superficielle ou tout au plus un instrument de recherche aveugle pour détecter des variations ayant une portée diagnostique si elle est tenue à l'écart de l'analyse psychopédagogique d'une matière donnée. Comme le fait remarquer Humphreys:

> Facets and the elements of facets are defined logically. Neither should be confused with factors. A facet is a logical dimension and its elements are the presence or absence of logically defined parts of that dimension. They must be manipulatable by the test constructer. They are not even very psychological as defined (1962, p. 480).

Les facettes qui peuvent être systématiquement introduites dans une épreuve de rendement n'échappent pas non plus à la critique formulée par Dur-

nin et Scandura (1973) à l'endroit des formes d'items. Plusieurs exemples de formes d'items trouvés dans la littérature apparaissent à première vue présenter des variations artificielles basées sur des caractéristiques perceptibles des questions ou des situations. Pour éviter ce genre d'inconvénient, notamment dans le cas des mathématiques, Durnin et Scandura ont proposé l'analyse de la matière en terme d'algorithmes. Dans le cadre de notre préoccupation, les facettes à introduire dans un test de mathématiques, par exemple, devraient être celles qui sont de nature à entraver l'usage ou l'application d'algorithmes de solution par des élèves en difficulté. Des études reliées à l'enseignement des mathématiques suggèrent des pistes valables à explorer dans cette direction. C'est ainsi que la retenue dans des problèmes d'addition s'est avérée une facette pertinente à introduire alors que la quantité de colonnes à additionner semble importer peu si ce n'est que d'exposer l'élève à un plus grand risque d'erreur d'inattention (Friend, 1979). Le choix des facettes à introduire dans un test apparaît donc comme un sujet de recherche et ce choix devrait être précédé d'une analyse psychopédagogique de la matière traitée dans ce test.

Une deuxième question qui émerge des exemples que nous avons donnés a trait au caractère diagnostique des profils que l'on peut établir en fractionnant l'ensemble des questions du test en divers sous-tests. La pertinence d'une ou de plusieurs facettes, même appuyée à priori par l'analyse psychopédagogique de la matière, doit être établie statistiquement pour constituer ce que certains auteurs appellent des catégories diagnostiques (Hively, Patterson et Page, 1968, par exemple). Dans une perspective de recherche pédagogique il ne suffit pas d'observer qu'un élève présente un profil différencié sur une ou plusieurs facettes du test pour déclarer que nous sommes en présence de catégories diagnostiques. Pour avoir une certaine signification, le phénomène doit être suffisamment répandu dans un échantillon d'élèves. On voit donc que la construction d'un test diagnostique d'après des facettes dictées par l'analyse psychopédagogique d'une matière doit être suivie d'une étude statistique des résultats obtenus avec un groupe d'élèves. C'est à l'égard de ce deuxième aspect du modèle de recherche pédagogique présenté dans ce document que sera envisagée l'analyse des composantes de la variance en rapport avec un test construit selon des facettes. Le chapitre IV

sera d'ailleurs consacré à la présentation des éléments essentiels de ce modèle de recherche et à la formulation des diverses hypothèses pertinentes au diagnostic pédagogique.

CHAPITRE III

RECHERCHES TRAITANT DE VARIATIONS INTRODUITES DANS DES QUESTIONS D'EXAMEN

Plusieurs études ont déjà démontré que des variations introduites dans des questions ou dans des problèmes peuvent être accompagnées de variations appréciables dans la difficulté de ces questions ou de ces problèmes. Qu'il suffise de mentionner les premières applications du schème à facettes multiples réalisées par Guttman et Schlesinger (1966) pour l'analyse des items d'un test d'analogies non verbales et l'analyse des items d'un test de compréhension mécanique. Dans un même ordre d'idées, Campbell a exploré en deux études distinctes l'influence de variations systématiquement introduites dans des tâches de classification de dessins géométriques (1961) et dans des problèmes d'analogies non verbales (1964). Scallon (1973) a repris substantiellement les tâches de classification de Campbell mais avec des contenus variés: figures géométriques simples, paires de lignes et groupements de lettres. Toutes ces études, caractérisées par la rédaction systématique des questions qui doivent composer un test, tendent à démontrer clairement qu'il est possible de manipuler la difficulté de ces questions avec des facettes appropriées. Cependant, les tests utilisés étant du domaine des aptitudes, toutes ces recherches ne semblent pas avoir d'impact immédiat sur le diagnostic pédagogique et sur l'enseignement correctif.

Pour ce qui est de l'étude de variations introduites dans des problèmes reliés à des matières académiques il existe une littérature abondante qui ne peut être recensée de façon exhaustive dans cet ouvrage. Seules les études les plus récentes, ayant trait de près ou de loin à un schème à facettes multiples et reliées à des matières académiques, seront présentées et discutées dans ce chapitre. Le but poursuivi en rapportant de telles études est de montrer l'intérêt que la recherche de variations à introduire dans des problèmes peut présenter pour le diagnostic pédagogique et l'enseignement correctif.

S'il est un domaine où il semble avoir été clairement établi que des variations introduites dans des situations de problèmes peuvent correspondre à des différences en taux de réussites c'est celui des additions et des soustractions de nombres naturels. On peut trouver à ce sujet des recherches qui remontent jusqu'à 1920. L'une des recherches les plus récentes en ce domaine est celle de Friend (1979) avec des problèmes d'addition posés à des élèves de niveau primaire résidant au Nicaragua. Les problèmes utilisés par Friend ont été choisis à même une banque considérable de problèmes d'addition de manière à faire varier les caractéristiques suivantes: le nombre de colonnes caractérisant les nombres à additionner, la quantité de nombres à additionner, la présence ou l'absence de zéros significatifs à manipuler et le nombre de retenues à effectuer. Ces variations, qui s'apparentent à des facettes, ont été envisagées dans le cadre d'une analyse de tâche c'est-à-dire qu'elles n'ont pas été inspirées de caractéristiques superficielles qui peuvent être observées dans des problèmes d'addition (Friend, 1979, p. 3). Des comparaisons appariées entre des problèmes ne différant qu'au point de vue d'une seule caractéristique ou facette ont amené les conclusions suivantes: le nombre de colonnes et la quantité de nombres à additionner n'ont pas été accompagnés d'un accroissement substantiel de la difficulté des problèmes. La retenue et la manipulation de zéros significatifs se sont avérés une source majeure de difficulté. Ces deux dernières variations présentent d'ailleurs des cas spécifiques que l'auteur a analysés avec détail. Les résultats de cette recherche peuvent avoir des répercussions sur l'élaboration des programmes reliés à l'enseignement des mathématiques élémentaires. D'après Friend, les variations qui se sont avérées significatives définissent des habiletés distinctes qu'il faudrait exercer en suivant une progression dans les types de problèmes à aborder d'un degré scolaire à un autre (1979, pp. 18-19).

A l'exemple des additions et des soustractions de nombres naturels, le domaine des fractions ordinaires a été exploré à plusieurs reprises. Il convient de signaler ici l'ouvrage d'Anna Bonboir (1970(b)), consacré à l'observation directe des procédés de solution utilisés par des élèves dans le but de déceler des fautes constantes commises dans la résolution de certains problèmes. En ce qui a trait à l'addition et à la soustraction de fractions

ordinaires, Anna Bonboir a présenté une typologie des fautes qui constitue un instrument d'analyse des plus précieux pour le diagnostic pédagogique, que celui-ci soit individuel ou collectif (voir pages 98 et suivantes). Cependant, l'approche suggérée par Bonboir relève du domaine de l'observation clinique ou du domaine de l'analyse détaillée des procédés de solution indiqués par des élèves à un examen administré collectivement. Des diverses études qui ont été menées au sujet des fractions l'étude d'Haseman (1981) est celle qui se rapproche le plus du domaine de la rédaction systématique d'une épreuve de rendement. Ce chercheur a mené une étude au sujet des difficultés associées aux fractions en se servant d'une épreuve construite selon trois facettes: 1) opérations sur les fractions, 2) forme de représentation des problèmes (diagrammes, symboles mathématiques, présentation verbale) et 3) formes opérationnelles. Cette dernière facette ne constitue pas un partage unique des problèmes et présente à notre avis un traitement non conventionnel de ce qui doit être entendu par facette d'après les textes publiés à ce sujet. Néanmoins, l'auteur a pu localiser chaque question de son épreuve dans un tableau de spécifications à trois dimensions. L'épreuve a été administrée à des groupes d'élèves dont l'âge variait entre 12 et 15 ans. Les résultats obtenus avec le groupe le plus faible ont été analysés en comparant le taux de réussites à des questions voisines dans le tableau de spécifications. Les interprétations avancées par Haseman sont beaucoup trop détaillées pour être résumées convenablement ici. En ce qui a trait à l'influence des facettes introduites, l'auteur rapporte les résultats suivants:

a) le mode de présentation des problèmes a été associé à des différences appréciables en difficulté de ces problèmes avec un taux relativement élevé d'échecs dans le cas des diagrammes;

b) dans le cas de la troisième facette (formes opérationnelles), les problèmes où il fallait trouver le tout à partir d'une fraction donnée se sont avérés beaucoup plus difficiles que les autres problèmes groupés selon cette facette.

D'autres résultats sont rapportés dans l'étude d'Haseman, notamment au niveau de comparaisons en taux de réussites entre plusieurs problèmes considérés deux-à-deux. Les conclusions et les recommandations de l'auteur touchent de près les moyens didactiques à mettre en oeuvre pour l'enseignement des mathématiques.

Dans le domaine des mathématiques on peut trouver une autre étude qui s'inspire d'un schème à facettes multiples mais qui cette fois a pour cible le domaine de la résolution de problèmes écrits. Depover (1979) a mené une étude sur l'effet de certaines variations pouvant être introduites dans des énoncés de problèmes d'arithmétique présentés verbalement. Ces variations sont: a) la présence ou l'absence de données superflues, b) la formulation de l'énoncé au présent ou au passé (temps de l'action), c) la formulation de la question au début ou à la fin de l'énoncé (place de la question) et d) l'explicitation ou non de certaines données contenues de manière implicite dans l'énoncé du problème. Les résultats obtenus avec des élèves de sixième année du primaire et traités au moyen de plusieurs analyses de la variance ont indiqué une augmentation significative de la difficulté des problèmes lorsqu'il y a présence de données superflues, lorsque la question posée est au début de l'énoncé et lorsqu'il y a explicitation de données implicites pour certains types de problèmes. Le temps du verbe n'est pas apparu comme une source significative de variation dans la difficulté des problèmes.

La description des problèmes d'arithmétique utilisés par Depover dans son étude indique que les énoncés ont été rédigés selon un schème à facettes multiples. Toutefois l'auteur n'a pas traité toutes les facettes qu'il a introduites dans une même analyse ce qui nous empêche d'apprécier l'influence de certaines combinaisons entre ces facettes. Les analyses rapportées par Depover visaient plutôt à étudier l'effet séparé de chacune des modifications introduites combinées avec le degré de difficulté des problèmes, le niveau de réussite des élèves à l'ensemble des problèmes ou encore avec la longueur de l'énoncé du problème. Il est intéressant de noter qu'aucune variation ou facette en interaction avec le niveau des élèves ne s'est avérée une source significative de variation en difficulté des problèmes. Enfin, les résultats d'analyse de la variance rapportés par l'auteur ne permet-

tent pas de retracer les composantes de la variance ce qui exclut pour l'instant la possibilité de tester le modèle de recherche pédagogique présenté dans cet ouvrage avec une étude qui contient, au départ, tous les éléments essentiels à ce modèle. Les tableaux d'analyse de la variance produits dans l'étude de Depover ne contiennent que des rapports F et il n'y a aucune mention quant à la nature, fixée ou aléatoire, des dimensions impliquées.

Les études impliquant de près ou de loin un schème à facettes multiples se font rares si nous quittons le domaine des mathématiques. Il en existe au moins une que l'on peut citer et qui est du domaine de l'orthographe d'usage en français. Garon, Ouellet, Savard et Dion (1978) ont mené une recherche sur l'accord du verbe avec son sujet en prenant des élèves de niveau secondaire. Un examen à choix de réponses a été construit de manière à couvrir un ensemble de vingt-cinq structures de phrases, avec quatre questions par structure. Pour chaque question les élèves devaient choisir l'une de cinq façons d'orthographier le verbe de la phrase parmi lesquelles se trouvait la bonne réponse. Les auteurs de la recherche ont compilé la fréquence de choix de chacune des mauvaises réponses et de la bonne réponse. Ainsi, le test utilisé dans cette étude s'apparente beaucoup plus au modèle des erreurs caractéristiques proposé par Guttman et Schlesinger (1967) qu'à un schème à facettes multiples des questions elles-mêmes. L'étude de la difficulté moyenne des questions se rapportant à une même structure de phrase a fourni cependant des indices pouvant guider la construction d'un test avec des facettes. Par exemple, la présence d'un écran entre le verbe et son sujet (comme dans: "l'odeur des sapins exhalait..."), la présence du pronom "nous" près du verbe ("ces camions nous transportaient..."), la présence de deux sujets placés avant ou après le verbe, l'emploi du pronom "on" et des expressions "c'est", "ce sont" etc. ont été identifiés dans l'étude comme étant des sources importantes de difficulté dans l'accord du verbe avec son sujet. Ces diverses sources de difficultés pourraient donc servir de facettes à introduire dans des structures de phrases pour composer une épreuve de rendement selon un schème à facettes multiples. D'autres sources de difficulté, que les auteurs ont identifiées, se prêteraient plutôt à une rédaction systématique des leurres pour des questions à choix de réponses. C'est le cas notamment de la marque du pluriel ("les lacs et les ruisseaux --dé-

gèles --dégèlents --dégèlent au printemps (Garon et alii, 1978, p. 23). L'approche utilisée par les auteurs de l'étude sur l'accord du verbe avec son sujet est sans contredit dans la poussée des tets diagnostiques. Cependant, plusieurs questions demeurent en suspens à savoir, par exemple, si les mêmes erreurs sont commises par les mêmes élèves. Cette interrogation est d'ailleurs formulée par les auteurs eux-mêmes (p. 21).

Les quelques études qui ont été rapportées dans ce chapître et les conclusions qui s'en dégagent suggèrent qu'il est possible de détecter ou d'identifier des symptômes pouvant accompagner la non maîtrise d'habiletés particulières. En aucun cas, cependant, les symptômes observés prescrivent par eux-mêmes les correctifs à apporter si ce n'est que de suggérer des pistes pour exercer les élèves dans des tâches spécifiques dans le but de prévenir ou de corriger les déficiences observées. Le modèle de recherche pédagogique proposé dans le présent ouvrage ne dépassera certes pas cette limite.

Les études qui ont été rapportées présentent certaines lacunes méthodologiques qu'il convient de signaler ici. Premièrement, ces études ne permettent pas d'observer l'influence de deux ou plusieurs variations combinées entre elles. Deuxièmement, il est rarement indiqué que l'influence d'une variation sur la difficulté des problèmes est la même pour tous les élèves ou se présente différemment d'un élève à un autre. Ces lacunes sont inhérentes à la nature unidimentionnelle du schème expérimental employé dans la plupart des études recensées. Pour ce qui est de l'étude de Depover (1979), qui est de nature multidimensionnelle au niveau des énoncés de problèmes écrits, l'emploi de l'analyse classique de la variance a permis d'éprouver des hypothèses tout en contrôlant plusieurs sources de variance. Cependant, l'analyse classique de la variance ne permet pas d'établir l'importance relative que chaque facette introduite dans une épreuve de rendement peut avoir sur la variance des résultats observés. Ce dernier aspect n'est pas à négliger dans une étude, même de nature exploratoire, qui a pour objet premier la détection de symptômes pouvant accompagner la non-maîtrise d'objectifs pédagogiques particuliers.

CHAPITRE IV

L'ANALYSE DES COMPOSANTES DE LA VARIANCE DANS UNE PERSPECTIVE DE DIAGNOSTIC PEDAGOGIQUE

Pour analyser les résultats obtenus avec un test construit selon des facettes et administré à un certain nombre d'élèves nous proposons comme outil statistique l'estimation et l'interprétation des composantes de la variance. Cette approche s'apparente à l'analyse classique de la variance à plusieurs points de vue mais s'en distingue nettement par l'esprit et par le but visé. L'analyse classique de la variance a pour objet premier d'éprouver, au moyen de rapports F, la signification statistique de l'influence d'un ou de plusieurs traitements introduits dans un schème expérimental. On peut ainsi détecter des sources significatives de variation parmi celles qui ont été introduites au moyen de traitements ou de facettes. En analyse classique de la variance les individus eux-mêmes, comme source identifiée de variation, présentent peu d'intérêt. Dans un schème à facettes multiples, l'étude des composantes de la variance permet d'évaluer l'importance relative de chacune des facettes pour expliquer les différences obtenues entre les résultats. Il est très important de souligner ici que pour l'étude des composantes de la variance les individus constituent eux-mêmes une dimension d'intérêt qui s'ajoute aux facettes du schème expérimental.

Au départ, tant pour l'analyse classique de la variance que pour l'estimation des composantes de la variance, le traitement statistique est le même. Ce n'est qu'à partir des carrés moyens que les deux modèles d'analyse statistique prennent des orientations différentes. Lorsque les composantes de la variance sont estimées, l'analyse peut être poussée plus loin du côté de la fiabilité des résultats obtenus en estimant, à partir de ces composantes, une variété de coefficients de fiabilité appelés plus spécifiquement coefficients de généralisabilité. Le modèle proposé dans cet ouvrage pour analyser les résultats obtenus à un test diagnostique s'inscrit donc dans le cadre général de la théorie de la généralisabilité. L'accent sera mis cependant sur les composantes de la variance et ce n'est qu'à l'occasion que seront

mentionnés quelques coefficients de généralisabilité qui peuvent être estimés en rapport avec le diagnostic pédagogique.

Ce chapitre traitera des caractéristiques les plus élémentaires de l'étude des composantes de la variance. Pour bien faire saisir l'esprit qui sous-tend ce procédé d'analyse, tout en évitant les développements mathématiques, un exemple fictif servira d'abord à établir les principales distinctions entre l'analyse classique de la variance et l'analyse des composantes de la variance. L'analyse proprement dite d'une épreuve construite selon des facettes sera ensuite illustrée à l'aide d'un autre exemple fictif. Pour terminer, un paradigme pouvant guider l'interprétation des composantes de la variance dans un schème à deux facettes sera présenté et discuté. Ce paradigme est au coeur du modèle de recherche pédagogique suggéré dans cet ouvrage puisqu'il identifie les composantes d'intérêt en rapport avec une typologie de symptômes accompagnant la non maîtrise d'un objectif dans un groupe d'élèves.

Comparaison entre l'analyse classique de la variance et l'étude des composantes de la variance.-

Voici un exemple pour illustrer la différence entre l'analyse classique de la variance et l'analyse des composantes de la variance. Supposons que 10 élèves aient réalisé une composition littéraire et que leurs copies aient été évaluées par 5 enseignants utilisés à titre de juges. Les notes attribuées par chaque enseignant à chaque copie apparaissent au tableau 1. Ce dispositif expérimental est de la famille des schèmes avec mesures répétées puisque les notes apparaissant dans chaque rangée du tableau 1 appartiennent à une même copie (ou à un même élève). Les notes apparaissant dans le corps principal du tableau 1 présentent une certaine variation ce qui, soit dit en passant, est tout-à-fait normal. On ne s'attendait pas à ce que tous les élèves produisent la même composition ni à ce que les juges aient une perception parfaitement univoque des copies des élèves. La variation entre les notes "élèves x juges" définit ce qui est appelé la variance totale en analyse de la variance. Cette variance totale est attribuée en partie aux différences entre les élèves et aux différences entre les juges. Les différences entre les élèves peuvent être inférées à partir des

notes moyennes obtenues par chaque élève (dernière colonne, à droite du tableau 1) et les différences entre les juges sont inférées à partir des notes moyennes attribuées à l'ensemble des élèves par chaque juge (dernière rangée du tableau 1). La variance totale ne saurait être expliquée

Tableau 1: notes obtenues par 10 élèves pour une composition littéraire évaluée par 5 enseignants.

élève	enseignant A	B	C	D	E	note moyenne par élève
1	4	5	2	0	0	(2,2)
2	6	9	6	2	4	(5,4)
3	8	15	8	7	7	(9,0)
4	3	14	12	9	8	(9,2)
5	4	6	1	0	1	(2,4)
6	9	10	10	6	1	(7,2)
7	5	10	3	5	5	(5,6)
8	2	10	5	4	1	(4,4)
9	4	6	4	2	4	(4,0)
10	4	8	2	3	1	(3,6)
note moyenne attribuée par juge	(4,9)	(9,3)	(5,3)	(3,8)	(3,2)	

uniquement par les différences entre les élèves et par les différences entre les juges. Il existe une certaine marge d'erreur qui, dans notre exemple, peut être attribuée à l'interaction élèves x juges confondue avec l'erreur expérimentale proprement dite. Les trois sources de variation ainsi identifiées apparaissent dans le tableau 2. En analyse classique de la variance les différences entre les individus importent peu. L'objet de l'étude sera plutôt de vérifier si les juges attribuent des notes différentes

et ce, de façon appréciable (les juges étant considérés ici comme produisant un effet de traitement). En supposant l'homogénéité des variances et des covariances et en supposant que l'interaction élèves x juges, confondue avec l'erreur expérimentale, peut servir de terme d'erreur, l'analyse de la variance servira à éprouver l'hypothèse que les juges proviennent d'une population de personnes qui perçoivent les compositions littéraires de la même façon et avec le même degré de sévérité. Il s'agit d'une hypothèse d'ordre purement statistique. Les résultats de cette analyse apparaissent au tableau 2. Le rapport F obtenu en rapport avec les différences entre juges (15,30) est significatif à $p < 0{,}001$ pour 4 et 36 degrés de liberté. On

Tableau 2: analyse de la variance des notes attribuées par cinq enseignants (juges) à dix compositions littéraires (élèves).

Source	S.C.	d.l.	C.M.	F
Entre les élèves	280,1	9	31,12	----
Entre les juges	228,2	4	57,05	15,30 [a]
Elèves x juges, e	134,2	36	3,73	
Total:	642,5	49	----	

a) significatif à $p < 0{,}001$

doit donc conclure que les juges ne proviennent pas d'une même population et qu'ils sont différents en sévérité comme le révèlent les notes moyennes attribuées par chacun d'eux à l'ensemble des élèves.

S'il est établi dans l'exemple qui vient d'être donné que les juges produisent des évaluations qui diffèrent significativement les unes des autres, l'analyse classique de la variance ne permet pas d'évaluer la contribution relative de ces différences à la variance totale. L'estimation des composantes de la variance nous permettra d'atteindre cet objectif. Le tableau 3 présente les estimés de ces composantes ainsi que la contribution

relative de chacune à la variance totale, contribution exprimée en pourcentage. Les formules qui apparaissent sous la rubrique "carrés moyens attendus" sont dérivées des lois de distributions d'échantillonnage d'une ou de plusieurs variances. Dans le cadre de ce document, il ne serait pas opportun de traiter en détail l'origine et le développement de ces diverses formules. Disons simplement que les valeurs attendues des carrés moyens font voir que chaque carré moyen observé correspond à une somme de composantes, certaines composantes étant pondérées par un nombre d'observations qui caractérise le dispositif expérimental employé. Ce sont les diverses expressions traduisant les carrés moyens attendus qui déterminent les algorithmes de calcul pour estimer la valeur de chaque composante. Ainsi, pour

Tableau 3: estimés des composantes de la variance et de leurs contributions relatives en pourcents en rapport avec l'exemple fictif des cinq juges qui ont évalué dix compositions littéraires.

Source	C.M.	carrés moyens attendus	estimés des composantes	%
Elèves (S)	31,12	$\sigma^2_{(sj,e)} + 5\,\sigma^2_{(s)}$	5,48	37,7
Juges (J)	57,05	$\sigma^2_{(sj,e)} + 10\,\sigma^2_{(j)}$	5,33	36,7
S x J, e	3,73	$\sigma^2_{(sj,e)}$	3,73	25,6
Total:	---	- - - -	14,54	100,0

estimer la valeur de la composante $\sigma^2_{(s)}$ on a suivi les étapes suivantes:

La différence entre le carré moyen "élèves" et le carré moyen "SxJ,e" équivaut à:

$$CM_{(s)} - CM_{(sj,e)} = \hat{\sigma}^2_{(sj,e)} + 5\,\hat{\sigma}^2_{(s)} - \hat{\sigma}^2_{(sj,e)} \quad (1)$$

$$CM_{(s)} - CM_{(sj,e)} \quad : \quad 5\,\hat{\sigma}^2_{(s)} \qquad (2)$$

Cette différence entre deux carrés moyens permet d'isoler la composante de la variance "élèves". La traduction des opérations (1) et (2) en valeurs numériques devient:

$$31,12 - 3,73 \quad : \quad 5\,\hat{\sigma}^2_{(s)} \qquad (3)$$

$$27,39 \quad : \quad 5\,\hat{\sigma}^2_{(s)} \qquad (4)$$

Puisqu'il nous faut maintenant isoler la composante $\hat{\sigma}^2_{(s)}$, on obtient, à partir de l'opération (4):

$$\hat{\sigma}^2_{(s)} \quad : \quad 27,39 \;/\; 5 \qquad (5)$$

$$\hat{\sigma}^2_{(s)} \quad : \quad 5,48 \qquad (6)$$

la valeur obtenue à l'opération (6) ayant été arrondie à deux décimales.

Une démarche analogue a été suivie pour estimer la composante "juges". Quant à la composante associée au terme d'erreur (sj,e) son estimé est le carré moyen "SxJ,e" tel quel. Les pourcentages apparaissant dans la dernière colonne du tableau 3 ont été obtenus en divisant chaque valeur estimée des composantes par la somme de ces valeurs estimées (qui est 14,54).

L'examen des valeurs estimées des composantes de la variance apparaissant au tableau 3 indique une contribution relativement élevée provenant des différences entre juges (36,7 %). Cette contribution est presqu'aussi importante que celle de la composante associée aux élèves (37,7 %). Cette première constatation nous amène à conclure que, d'après les estimations faites à partir d'un dispositif expérimental impliquant 10 élèves et 5 juges, les différences observées entre les notes attribuées à des compositions littéraires seraient tout aussi bien expliquées par les différences entre juges que par les différences en compétence chez les élèves. La contribution relative associée au terme d'erreur (25,6 %) ne doit pas être négligée non plus. Ces quelques remarques font voir comment l'étude des composantes de la variance peut nous aider à apprécier un dispositif expérimental choisi.

L'interprétation des composantes de la variance constitue une première étape de l'application de la théorie de la généralisabilité. Si l'évaluation de compositions littéraires par des juges nous préoccupe, la théorie de la généralisabilité offre des raffinements qui vont au-delà de l'étude proprement dite des composantes de la variance. Dans le cadre de cette théorie, l'impact de certaines composantes sur la fiabilité des notes obtenues par des élèves, lorsque leurs compositions sont évaluées par des juges, peut être établi de plusieurs façons. Les composantes d'intérêt sont dans ce cas la composante "élèves" et la composante "élèves x juges, e".

La fiabilité des notes que reçoivent les élèves de la part d'un seul juge:

$$\frac{5,48}{5,48 + 3,73} = 0,59$$

La fiabilité des notes moyennes que recevraient les élèves de la part des cinq juges:

$$\frac{5,48}{5,48 + 3,73/5} = 0,88$$

Si le contexte de l'évaluation des compositions littéraires ne change pas il serait fortement souhaitable d'utiliser une note tenant compte des évaluations faites par les cinq juges plutôt que la note attribuée par un seul juge. La fiabilité de ce dernier type de résultat est 0,59 alors qu'elle atteint 0,88 avec cinq juges.

Dans les pages qui précèdent, un parallèle a été tracé entre l'analyse classique de la variance et l'analyse des composantes de la variance. Les nombreuses technicalités qui caractérisent les deux approches ont été évitées le plus possible. Le lecteur intéressé à la théorie de la généralisabilité trouvera dans plusieurs ouvrages spécialisés sur le sujet des informations beaucoup plus techniques (à titre d'exemples: Cronbach et alii (1972), Cardinet, Tourneur et Allal (1976), Brennan et Kane (1979), Tourneur et Cardinet (1980).

Pour ce qui est du diagnostic pédagogique et plus particulièrement d'un test construit selon des facettes, l'analyse statistique au moyen de l'estimation des composantes de la variance permettra de détecter les facettes qui contribuent le plus à la variance totale et qui de ce fait sont associés à des variations importantes dans la difficulté des questions du test. Mais cette analyse ne se limite pas aux seules facettes du test. La plupart des schèmes utilisés permettront d'identifier plusieurs sources de variation au niveau des interactions entre ces facettes d'une part et au niveau des interactions entre chaque facette et les sujets d'autre part. L'étude des résultats à un test diagnostique construit selon des facettes au moyen des composantes de la variance révèlera ainsi plusieurs nuances qui découlent d'éléments rationnels qui seront présentés à la fin de ce chapitre. L'exemple des compositions littéraires n'avait pas de portée diagnostique en lui-même. Avant de présenter le modèle proprement dit de recherche pédagogique qui est suggéré dans cet ouvrage il conviendrait d'étudier maintenant le cas d'un examen construit selon deux facettes et qui a été administré à un groupe d'élèves. Ce deuxième exemple, qui présente un cas fictif, a été conçu pour illustrer les nuances qui s'imposent dans la recherche de profils pouvant correspondre à des catégories diagnostiques.

Analyse d'un examen fictif rédigé selon deux facettes.-

Supposons qu'une épreuve de 9 questions ait été rédigée selon un schème à deux facettes croisées qui seront désignées par les lettres A et B. L'épreuve a été administrée à 10 élèves. Les résultats obtenus à la suite d'une correction dichotomique des questions (succès: 1, échec: 0) apparaissent au tableau 4. Un coup d'oeil rapide sur ces résultats révèle une certaine variation qu'il faudra essayer d'expliquer au moyen de sources identifiées: les élèves, les facettes de l'épreuve et les interactions.

Pour apprécier l'influence de chaque facette, il faut réduire les données de base. Cet objectif peut être atteint en partie en calculant le nombre moyen de réussites aux questions groupées sous chaque facette. Ainsi, l'influence de la facette A s'observe en calculant un nombre moyen de réussites aux questions a_1, aux questions a_2 et aux questions a_3. A par-

tir des résultats présentés au tableau 4 on obtient les moyennes suivantes:

Facette A: a_1 : 1,2 a_2 : 1,6 a_3 : 1,7

Facette B: b_1 : 2,6 b_2 : 1,2 b_3 : 0,7

A première vue, la facette B semble correspondre à des différences en difficulté des questions de façon plus marquée que la facette A. D'ailleurs, une analyse classique de la variance révèlerait une différence non significative entre les questions groupées sous la facette A et une différence significative entre les questions groupées sous la facette B (à un niveau de probabilité inférieur à 0,001 et en supposant l'homogénéité des variances et des covariances car les données de base proviennent d'un exemple fictif).

Tableau 4 : succès (ou échecs) obtenus par 10 étudiants fictifs à une épreuve composée de 9 questions rédigées selon deux facettes croisées A et B

élève	a_1			a_2			a_3		
	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
(1)	0	0	0	1	1	1	1	0	0
(2)	1	1	0	1	0	0	0	0	0
(3)	1	0	0	1	1	0	1	1	1
(4)	1	0	0	1	1	1	1	0	0
(5)	1	1	1	1	1	0	1	0	1
(6)	0	0	0	1	0	0	1	0	0
(7)	0	0	0	1	0	0	1	1	0
(8)	1	1	0	1	0	0	1	1	1
(9)	1	1	0	1	0	0	1	0	0
(10)	1	0	0	1	0	0	1	1	1

A la suite de cette analyse sommaire on serait alors tenté de conclure que seule la facette B présente les caractéristiques de ce qui pourrait être une catégorie diagnostique. Les élèves, comme individus, ont été

ignorés jusqu'ici. Pour étudier le comportement des facettes chez les élèves il est nécessaire d'observer les scores profils de chacun de ceux-ci, ce qui constituerait une tâche des plus fastidieuses avec un grand nombre d'élèves. Le tableau 5 présente les profils de chaque élève, exprimés en nombres de questions réussies, aux sous-tests formés d'après la facette A d'une part et aux sous-tests formés d'après la facette B d'autre part. L'examen des deux types de profils suggère des interprétations qui apparaissent dans le corps du tableau 5. Les profils des élèves révèlent certaines par-

Tableau 5 : profils des élèves en regard de chacune des facettes à partir des résultats apparaissant au tableau 4.

FACETTE	sous-test	élève (1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	Tot.
A	a_1	0	2	1	1	3	0	0	2	2	1	12
	a_2	3	1	2	3	2	1	1	1	1	1	16
	a_3	1	0	3	1	2	1	2	3	1	3	17
		Interprétation: la difficulté des sous-tests varie d'un élève à l'autre; pour l'ensemble des élèves les sous-tests sont sensiblement de même difficulté.										
B	b_1	2	2	3	3	3	2	2	3	3	3	26
	b_2	1	1	2	1	2	0	1	2	1	1	12
	b_3	1	0	1	1	2	0	0	1	0	1	7
		Interprétation: la difficulté des sous-tests varie moins d'un élève à l'autre; le sous-test b_1 est apparu le plus facile pour tous les élèves; pour l'ensemble des élèves les sous-tests sont plus différenciés en difficulté										

ticularités que masquait l'interprétation des moyennes globales de réussites aux questions groupées sous chaque facette. Les sous-tests formés d'après la facette A, qui apparaissaient quelque peu voisins en difficulté pour l'ensemble des élèves, se présentent différemment d'un élève à l'autre. Pour

l'élève (1) les questions a_2 ont été mieux réussies que les questions a_1 et a_3; pour l'élève (2), ce sont les questions a_1 qui sont les mieux réussies mais de façon moins marquée; enfin, pour l'élève (10), les questions a_3 ont été les mieux réussies. La différenciation entre les sous-tests a_1, a_2 et a_3 varie en direction d'un élève à l'autre ce qui constitue la marque distinctive d'une interaction possible entre la facette A et les élèves. La facette B présente par contre un comportement différent de celui de la facette A. Les questions b_1 ont manifestement été mieux réussies que les questions b_2 et b_3, et les questions b_2 ont été à leur tour mieux réussies que les questions b_3. L'ordre de difficulté croissante de b_1 à b_3, entre les sous-tests, s'observe pour chaque profil d'élève. Ce phénomène suggère une absence d'interaction entre la facette B et les élèves ce qui signifierait vraisemblablement que la facette B introduit des différences en difficulté des questions et ce indépendemment des élèves. On voit donc que l'examen des profils d'élèves en regard de chaque facette fournit déjà des indications plus nuancées que l'examen des moyennes globales aux divers sous-tests dans la recherche de catégories diagnostiques. La facette A, qui à première vue, présentait peu d'intérêt du point de vue diagnostique, présente maintenant une toute autre image: l'interaction que cette facette entretient avec les élèves indique que nous sommes en présence d'un moyen efficace pour accroître la difficulté des questions mais que ce moyen n'a pas le même effet chez tous les élèves. Le phénomène serait alors particulier à chaque élève ou à certains élèves et une étude plus poussée devrait être entreprise pour expliquer l'interaction entre la facette d'intérêt et les élèves en identifiant chez ces derniers des caractéristiques individuelles comme les aptitudes, la motivation ou le degré de maîtrise d'habiletés requises à la poursuite de l'objectif mesuré par le test. En ce qui a trait à la facette B, ses éléments b_1, b_2 et b_3 correspondent à une difficulté accrue et ce de façon presqu'identique pour tous les élèves. L'effet de cette facette ne dépendrait donc pas des élèves ou de certains élèves. La facette B correspondrait ainsi à des symptômes de non maîtrise de l'objectif mesuré par le test, symptômes qui présentent un caractère généralisé à tous les élèves examinés. Un tel événement invite à chercher du côté de la pédagogie ou de la didactique ou encore du côté des théories de l'apprentis-

sage pour expliquer pourquoi les sous-tests b_1, b_2 et b_3 présentent un ordre croissant de difficulté, quels que soient les élèves.

L'examen des profils d'élèves, combiné à l'étude des moyennes de réussites aux divers sous-tests, constitue une approche intuitive à la recherche de catégories diagnostiques. Avec un grand nombre d'élèves, les moyens proposés jusqu'à maintenant peuvent s'avérer très onéreux. De plus, les profils établis dans l'exemple choisi présentent des caractères visibles à l'oeil nu ce qui ne sera pas toujours le cas avec des données réelles. Signalons enfin que l'analyse menée jusqu'ici demeure incomplète à plusieurs points de vue. Qu'en est-il de l'influence réciproque que pourraient avoir l'une sur l'autre les facettes A et B pour produire un effet combiné qu'on appelle interaction? Sachant que les facettes A et B constituent des catégories diagnostiques, chacune à sa façon bien sûr, quelle importance doit-on leur accorder à titre de sources de variation des résultats obtenus? Sachant, par expérience, qu'aucune mesure n'est infaillible, quelle confiance doit-on accorder à tous les aspects descriptifs présentés plus haut? Cette dernière question est fondamentale puisqu'elle touche de près l'erreur expérimentale ayant pu contaminer le dispositif utilisé dans l'expérience. Les réponses à toutes ces questions ne peuvent être obtenues avec des procédés intuitifs. C'est pourquoi l'analyse des composantes de la variance a été proposée comme complément à la construction d'une épreuve de rendement selon des facettes.

Le tableau 6 présente les estimés des composantes de la variance en rapport avec l'exemple fictif décrit au tableau 4 ainsi que les contributions relatives de ces composantes, exprimées en pourcentages de la somme des composantes. Les sources de variation identifiées, les sommes de carrés (S.C.), les degrés de liberté (d.l.) et les carrés moyens (C.M.) sont les éléments de départ d'une analyse classique de la variance. Les estimés des composantes proviennent d'opérations algébriques effectuées sur les carrés moyens à partir de formules exprimant les valeurs attendues de ces carrés moyens. Ces formules dépendent du caractère fixé ou aléatoire des dimensions de l'analyse. Pour l'exemple fictif choisi, les deux facettes de l'épreuve ont été considérées comme fixées et les élèves (ou sujets) comme étant une dimension aléatoire simple (en supposant que les élèves ont été échantillonnés à

partir d'une population infinie d'élèves). C'est donc dire qu'au niveau des conclusions, nous n'envisagerons pas la possibilité d'introduire les facettes A et B autrement que par les éléments de ces facettes qui ont été effectivement utilisés. En considérant les facettes d'une épreuve comme fixées (ce qui sera toujours le cas dans ce document) les conclusions sont formulées dans la perspective de réutiliser la même épreuve. En considérant

Tableau 6 : estimés des composantes de la variance en rapport avec les résultats fictifs à une épreuve de 9 questions rédigées selon deux facettes croisées A et B

Source[a]	S.C.	d.l.	C.M.	$\hat{\sigma}^2$	%
Entre les élèves	2,5	9	0,277	0,0308	7,3
Facette A	0,467	2	0,234	0[b]	----
Facette B	6,467	2	3,234	0,1062	25,2
Elèves X A	6,2	18	0,344	0,1147	27,2
Elèves X B	0,866	18	0,048	0,0160	3,8
A X B	0,466	4	0,117	0[b]	----
El. x A x B,e	5,534	36	0,154	0,154	36,5
Total:	22,5	89	-----	0,4216	100,00

a: les élèves ont été considérés comme aléatoires, les facettes A et B, fixées.
b: la valeur zéro a été substituée à une valeur négative.

les élèves comme une dimension aléatoire les conclusions tirées s'étendent à des élèves autres que ceux qui ont été utilisés au moment de l'expérience.

Les formules mathématiques utilisées pour estimer les composantes de la variance à partir des carrés moyens observés sont reproduites en appendice.

L'interprétation des contributions relatives à la variance totale, en regard des diverses sources de variation apparaissant au tableau 6, re-

joint sur le plan descriptif ce qui a été observé auparavant en examinant les profils A et B de chaque élève: de toutes les sources identifiées, sauf l'erreur expérimentale, la facette B apparaît comme étant la plus importante avec 25,2 % de contribution relative à la variance totale; la facette A à elle seule serait une source négligeable de variation avec une composante négative à laquelle on a donné la valeur zéro tel que suggéré en maints ouvrages (par exemple: Kirk, 1968, p. 135; Cronbach et alii, 1972, p. 57); enfin, l'interaction déjà soupçonnée entre la facette A et les élèves émerge de façon claire avec 27,2 % de contribution relative.

L'analyse des composantes de la variance apporte des éléments de réponse à d'autres questions qui ont été formulées auparavant. Les effets de la facette B, laquelle s'est avérée une source importante de variation, seraient indépendants des élèves et indépendants de la facette A. Les interactions "élèves x B" et "A x B" ont produit des estimés de valeur négligeable.

Il reste un dernier point, et non le moindre, sur lequel il faut attirer l'attention. Il s'agit de la composante d'erreur dont la contribution relative à la variance totale est 36,5 %. Ce pourcentage, qui est le plus élevé de tous, indique sans aucune ambiguïté que le dispositif expérimental utilisé a subi l'influence de sources de variation non identifiées. En l'absence de données précises au sujet du contenu de l'épreuve et au sujet du contexte dans lequel celle-ci a été administrée, il est pratiquement impossible de suggérer des améliorations de nature qualitative. Sur le plan quantitatif, on peut lever certains soupçons. La nature dichotomique des résultats analysés fait que la variation totale peut être inférieure à la variation qui serait observée dans le cas d'une variable continue. Le petit nombre de questions qui composent l'épreuve doit également être pris en considération. On pourrait donc augmenter le nombre de questions correspondant à chaque cellule du tableau de spécifications des facettes. Les élèves choisis pour répondre aux questions de l'épreuve ne présentent pas non plus beaucoup de différences. Le tableau 6 indique une contribution relative de 7,3 % pour ce qui est des différences entre élèves. Administrer le test à un plus grand nombre d'élèves qui, par surcroît, présenteraient des différences plus marquées en compétence ou en degré de maîtrise de l'objectif, pourrait s'avérer une avenue intéressante à explorer.

Les divers commentaires formulés à l'endroit du dispositif expérimental qui caractérise l'exemple fictif qui vient d'être décrit illustrent le genre de conclusion que l'on peut tirer à partir d'une analyse de composantes de la variance et ce, dans l'esprit même d'une étude de généralisabilité. Ce genre d'étude, appelée "étude G" par Cronbach et alii (1972), a un caractère exploratoire et peut fournir des indications utiles à des recherches ultérieures. Il devient donc possible d'améliorer graduellement, c'est-à-dire par études successives, un dispositif expérimental donné. L'établissement de profils diagnostiques pour une épreuve de rendement destinés à mesurer un objectif important est en soi un objet de recherche qui peut être poursuivi par une ou plusieurs études de généralisabilité.

Avant de terminer cette section il apparaît important de préciser un point qui a été délibérément laissé en suspens jusqu'à maintenant. Les composantes de la variance, comme celles qui ont été estimées pour l'exemple fictif d'une épreuve à deux facettes, sont en étroite relation avec la notion de fiabilité en théorie de la généralisabilité. Lorsqu'il est question de fiabilité il faut toujours préciser de quels résultats nous voulons parler: du résultat que chaque élève obtient à une seule question? du résultat que chaque élève obtient à trois questions d'un sous-test? du résultat que chaque élève obtient à toutes les questions du test? ou encore, du résultat moyen obtenu par l'ensemble des élèves à chaque sous-test? Cette précision doit toujours être explicitée avec le plus grand soin. Par exemple, la composante d'erreur estimée pour le résultat obtenu par un élève à une seule question de l'épreuve est 0,154 (voir dans la colonne des estimés de composantes de la variance du tableau 6); cette même composante d'erreur estimée pour le résultat obtenu par un élève aux neuf questions de l'épreuve serait 0,154/9 soit 0,017. Dans cette optique, certaines des composantes du tableau 6 seraient ainsi pondérées par un nombre d'observations déterminé par les lois de distribution d'échantillonnages. La contribution relative de chaque composante sera donc modifiée selon que le résultat cible de chaque élève est le nombre de réussites à chaque sous-test ou à l'ensemble des questions de l'épreuve. Dans une étude de type D (étude de décision, d'après la nomenclature de Cronbach et alii (1972)), il s'agira alors

de déterminer si le score d'un élève à trois questions par exemple (ce qui correspond aux profils déjà identifiés) peut être interprété de façon fiable. Il s'agit là d'un objet de recherche distinct de celui qui caractérise une étude de type G.

Le point que nous venons de soulever nous amène à préciser un élément qui apparaissait difficile à établir au début de ce document: le modèle d'analyse proposé pour l'étude d'un test construit selon des facettes se situe donc au niveau d'une étude de généralisabilité, c'est-à-dire d'une étude de type G. L'objet d'une telle étude est de décrire, au moyen de composantes estimées de la variance, les effets des principales sources de variation introduites dans une épreuve de rendement et d'inférer éventuellement l'existence de catégories diagnostiques. En d'autres mots, il s'agit d'une recherche de symptômes pouvant accompagner la non maîtrise d'un objectif dans un groupe d'élèves. Cette démarche n'est certes pas étrangère à l'estimation de la fiabilité des profils à être établis si l'épreuve devait être utilisée telle quelle dans un contexte pédagogique où des décisions de nature corrective seront prises. Ce sujet de matière exigerait à lui seul un traitement assez élaboré qui dépasserait de beaucoup les limites du présent ouvrage. C'est pourquoi le modèle de recherche pédagogique qui est suggéré ici s'appuie sur une première étape de l'application de la théorie de la généralisabilité qui est l'analyse des composantes de la variance.

Les composantes d'intérêt pour l'analyse d'un test construit selon des facettes dans une perspective de diagnostic pédagogique.-

Nous avons vu que la construction systématique d'un test selon un schème à facettes s'apparente à l'élaboration d'un schème multifactoriel d'analyse de la variance. Toutefois, dans l'esprit de la théorie de la généralisabilité, les individus ou personnes s'ajoutent aux dimensions du schème pour devenir une autre source de variation d'intérêt. Cronbach et alii (1972) ont d'ailleurs tenu à séparer, dans leur terminologie, les personnes des facettes ou dimensions du schème expérimental proprement dit. L'administration d'un test,construit selon une, deux ou trois facettes, à n personnes, nous conduira habituellement à un schème avec mesures répétées sur toutes les fa-

cettes; ce schème sera aussi habituellement caractérisé par la présence d'une seule observation par cellule ou par combinaison de dimensions. Dans certains cas, les questions elles-mêmes qui composent le test deviendront une autre dimension. Il devient donc extrêmement important de pouvoir identifier toutes les sources de variation qui entrent dans un dispositif expérimental donné. Les quelques exemples qui suivent donneront un aperçu de ce qui doit être entendu par l'identification des sources de variation. Le lecteur intéressé à cette question trouvera des informations plus complètes dans des manuels d'analyse de la variance.

Supposons que l'on ait des problèmes de pourcentage à faire résoudre par des élèves et que nous disposions d'un échantillon représentatif de dix énoncés écrits de ces problèmes. On veut faire varier la position de la question dans l'énoncé, de la façon suivante: pour les dix énoncés, on formule d'abord la question au début et pour les mêmes dix énoncés on formule la question à la fin de l'énoncé. On obtient ainsi vingt questions à poser que l'on disposera dans une séquence selon un procédé d'assignation au hasard. Cependant, il faut comprendre que chaque énoncé apparaît deux fois dans le test. Si le test est administré à un certain nombre d'élèves choisis au hasard on obtient les sources suivantes de variation:

la position de la question (facette introduite dans le test)		a
les 10 problèmes rédigés		b
les élèves		p
l'interaction	a x b	
l'interaction	p x a	
l'interaction	p x b	

et l'interaction p x a x b, confondue avec l'erreur expérimentale puisqu'il n'y a qu'une seule observation par cellule du tableau de spécifications des facettes. Les dimensions de ce schème sont alors complètement croisées.

Cependant, si pour étudier l'influence de la position de la question dans l'énoncé d'un problème il fallait utiliser deux échantillons de problèmes

(ce qui, soit dit en passant, constituerait un dispositif plus approprié que le précédent) la dimension "problèmes" serait alors nichée sous la dimension "position de la question" puisque sous chaque élément de la facette du test il y aurait un ensemble différent de problèmes. On aurait donc au total 20 problèmes distincts à rédiger et non pas 10 comme précédemment. Dans ce deuxième cas, les diverses sources de variation seraient:

la position de la question dans l'énoncé	a
les problèmes nichés sous la facette "position de la question"	b:a
les élèves	p
l'interaction p x a	

l'interaction p x b:a, confondue avec l'erreur expérimentale. Il s'agit alors d'un schème partiellement niché.

Traiter de tous les schèmes possibles n'est pas le but poursuivi dans ce document. Le premier et le deuxième exemples font voir que dans ces deux cas la source d'intérêt du point de vue diagnostique, c'est-à-dire la position de la question dans l'énoncé, peut être isolée. Le premier exemple souffre de certaines lacunes méthodologiques que présente la répétition des mêmes problèmes à résoudre dans une épreuve de rendement. Le second exemple offre un schème expérimental qui évite cette lacune mais qui ne permettra pas d'étudier l'effet possible de l'interaction p x b (élèves x problèmes). Ces deux exemples font donc voir que l'identification des sources de variation dépend du schème expérimental utilisé et que cette étape doit être franchie avant même que l'expérience proprement dite soit réalisée si l'on veut s'assurer au départ que toutes les sources de variation qui présentent un intérêt particulier pourront être isolées les unes des autres.

Lorsque les sources de variation qui entrent dans un dispositif expérimental choisi ont été correctement identifiées quelles sont les conclusions que l'on peut tirer de la valeur de composante de la variance qui est associée à chacune d'elles? La réponse à une telle question n'est pas simple car il y a plusieurs éléments à surveiller. Dans cette dernière partie de ce chapitre les aspects rationnels en rapport avec cette question

seront présentés et discutés dans le cadre du diagnostic pédagogique.

La première composante de la variance à examiner de près est celle qui est associée à l'interaction du plus haut degré confondue avec l'erreur. Dans le cas du premier exemple cité précédemment il s'agit du terme "p x a x b" et dans le cas du deuxième exemple, il s'agit du terme "p x b:a". Si la valeur estimée de cette composante (ou sa contribution relative en pourcent) est relativement élevée en comparaison avec la valeur estimée des autres composantes, celà signifie que les variations observées entre les données de base sont principalement dues à de l'erreur expérimentale. Celà signifie également que d'autres sources de variation non contrôlées ont exercé une certaine influence. Pour faire une telle affirmation il faut toujours supposer que l'interaction du plus haut degré entre les dimensions du schème expérimental, qu'on n'a pu isoler de l'erreur proprement dite, est négligeable au point d'être considérée comme nulle. Ceci revient à dire que l'on doit supposer que le terme "p x a x b" (ou "p x b:a" selon l'exemple) est un estimé du terme d'erreur. C'est l'un des principaux inconvénients du dispositif expérimental qui ne comporte qu'une seule observation par cellule.

Outre la composante associée à l'erreur expérimentale, les composantes d'intérêt sont sans contredit celles qui sont associées à chacune des facettes introduites dans la construction du test diagnostique. Dans notre premier exemple il n'y en a qu'une; c'est celle qui est associée à la position de la question dans l'énoncé d'un problème. Si la valeur de cette composante est relativement élevée celà indique que la position de la question est associée à une différence appréciable dans la difficulté de résolution des problèmes. Un retour sur les donnés de base indiquerait alors dans quelle direction s'établit cette différence à savoir pour quelle position dans l'énoncé (au début ou à la fin) les problèmes se sont avérés les moins réussis ou les plus difficiles.

L'influence de chaque facette peut ainsi être évaluée à titre d'effet principal. Toutefois, l'analyse ne s'arrête pas là. Une facette à elle seule peut ne pas amener de différence en difficulté des questions ou problèmes mais produire des différences non négligeables lorsque combinée avec d'autres facettes: c'est au niveau des interactions entre cette fa-

cette et d'autres facettes que l'on peut déceler ce genre de phénomène. L'interaction entre deux facettes indique que la difficulté introduite au moyen de l'une de ces facettes varie en importance lorsque les éléments de l'autre facette sont considérés à tour de rôle. L'interprétation d'une interaction est une entreprise fort délicate et on trouvera dans plusieurs ouvrages de statistique des exemples de procédés graphiques à utiliser pour visualiser ce qui se passe lorsque deux ou plusieurs facettes ou dimensions semblent s'influencer mutuellement.

Il existe cependant un type d'interaction qui a été négligé dans la littérature en analyse classique de la variance et qui peut s'avérer des plus pertinentes à étudier en matière de diagnostic pédagogique. Il s'agit de l'interaction entre une facette du test et les sujets. Si ce type d'interaction correspond à une contribution relativement importante à la variance totale on peut supposer que la facette d'intérêt est associée à des différences en difficulté qui ne sont pas dans le même sens ou qui n'ont pas la même importance pour tous les sujets. Une facette qui amènerait des différences en difficulté, quels que soient les sujets, présenterait une interaction négligeable avec la dimension "sujets". La portée diagnostique de l'interaction "facette x sujets" se présente d'elle-même. Si l'introduction d'une facette augmente d'une manière substantielle la difficulté des problèmes et ce, quels que soient les sujets (absence d'interaction sujets x facette),nous sommes alors en présence d'un symptôme de nature collective c'est-à-dire d'un symptôme qui est généralisable à la plupart des sujets. Si la difficulté introduite n'est pas la même (ou ne va pas dans le même sens) pour tous les sujets nous sommes alors en présence d'un symptôme de nature idiosyncratique. Au niveau du correctif la recherche pourra être orientée du côté de la révision de programmes, de la psycho-pédagogie ou de la didactique si l'effet d'une facette est généralisable à la majorité des élèves. Par contre, si l'effet d'une facette n'est pas le même pour tous les élèves (interaction facette x sujets) il faudra alors orienter la recherche du côté des facteurs individuels pouvant expliquer cette disparité (niveau d'intelligence, aptitudes spécifiques, styles d'apprentissage, motivation, maîtrise d'habiletés requises à l'objectif cible etc.). Le modèle de recherche pédagogique qui est proposé dans ce document présente donc deux débouchés possibles au ni-

veau du correctif selon l'interprétation qui peut être faite de l'effet d'une facette et de l'interaction entre cette facette et les sujets.

En résumé, la composante d'erreur, les composantes associées aux diverses facettes du test prises isolément ou en interaction et les composantes d'interaction avec les sujets constituent les principales cibles de l'analyse d'un test construit selon des facettes à des fins de diagnostic pédagogique. L'interprétation de tous ces éléments suit d'ailleurs d'assez près les règles énoncées dans plusieurs manuels traitant de l'analyse de la variance. Mais l'étude des composantes de la variance ne vise pas à rechercher des effets significatifs. Son objectif principal est de détecter les sources qui contribuent principalement à expliquer les différences observées entre les résultats d'un dispositif expérimental donné.

Le tableau 7 permet de visualiser, tout en les résumant, les principaux éléments rationnels du modèle de recherche pédagogique suggéré dans ce document. Ce tableau présente cinq patrons possibles de composantes de variance pour un schème du type sujets x facette A x facette B (équivalent au schème p x a x b selon les symboles rencontrés dans plusieurs ouvrages). L'énumération de ces patrons n'est pas exhaustive car seules les situations les plus caractéristiques ont été signalées. On remarquera que les patrons III, IV et V indiquent trois façons de détecter une ou des catégories diagnostiques. Ces patrons doivent être discutés avec beaucoup de réserve car seule la facette A a été choisie comme cible dans un contexte où elle apparaît relativement importante à titre d'effet principal. Il faut comprendre que d'autres cas possibles peuvent émerger au-delà des patrons III, IV et V, chacun de ces cas exigeant une interprétation plus nuancée.

Ce chapitre avait pour but de faire voir, théoriquement du moins, que l'analyse des composantes de la variance peut s'avérer un outil fort commode de recherche pédagogique pour interpréter des difficultés à réussir un ensemble de questions ou de problèmes. La construction d'un test selon un schème à facettes multiples et l'analyse statistique des résultats qui en découlent, lorsque les élèves sont identifiés comme une dimension de l'étude, sont apparues pour l'instant compatibles avec l'analyse des composantes de la variance et ce dans l'esprit même de la théorie de la généralisabilité. Avec des exemples fictifs nous avons essayé d'illustrer le genre de symptôme (in-

Tableau 7 : présentation de quelques patrons possibles de contributions relatives des composantes de la variance reliées à une facette A, dans un schème S(sujets) x A x B avec une observation par cellule, et interprétation du point de vue diagnostique.

source	PATRON I	II	III	IV	V
Sujets: **S**	○●	●	○●	○●	○●
Facette: **A**	○	○	●	●	●
Facette: **B**				○	
AxB	○	○	○	●	○
SxA	○	○	○	○	●
SxB					
SAB,e	●	○	○	○	○

Légende: ○ :contribution négligeable ● :contrib. importante
○● :contribution négligeable ou importante

I N T E R P R É T A T I O N:

I sources de variation non contrôlées et/ou interaction SxAxB non nulle;

II bonne différenciation entre les sujets mais les facettes introduites ne sont pas associées à des variations importantes en difficulté;

III la facette A correspond à des variations importantes en difficulté, pour toute condition de B et de la même façon pour l'ensemble des sujets; — FACETTE A: CATEGORIE DIAGNOSTIQUE

IV A correspond à des variations importantes en difficulté mais seulement en présence de certaines conditions de B ; — COMBINAISON 'AB': CATEGORIE DIAGNOSTIQUE

V A correspond à des variations importantes en difficulté mais qui ne vont pas dans le même sens pour tous les sujets — FACETTE A: CATEGORIE DIAGNOSTIQUE NON GENERALISABLE A TOUS LES SUJETS

dividuel ou collectif) qui peut être détecté devant la non-maîtrise d'un objectif pédagogique. La détection peut se faire intuitivement dans certains cas mais l'examen des composantes de la variance peut révéler des phénomènes qui ne sont pas toujours visibles à l'oeil nu.

CONCLUSION A LA PREMIERE PARTIE

Le modèle de recherche pédagogique présenté dans cette première partie de notre ouvrage s'inscrit dans l'étude des variations à introduire dans les questions, dans les situations ou dans les problèmes qui composent une épreuve de rendement centrée sur un objectif et destinée au diagnostic pédagogique. Dans une perspective de construction systématique d'une épreuve ou d'un test diagnostique, les variations à introduire constituent des facettes. Chacune de ces variations peut être accompagnée de différences plus ou moins marquées dans la difficulté des questions. L'objet même du modèle de recherche pédagogique proposé dans cet ouvrage est la recherche de variations ou de facettes qui amènent des différences importantes en taux de réussites aux questions d'un test destiné au diagnostic des difficultés d'apprentissage. De telles variations correspondent à ce que nous avons appelé les symptômes de la non-maîtrise de l'objectif pédagogique mesuré par le test. En regard de chaque facette introduite comme dimension de l'étude, à laquelle s'ajoute la dimension "élèves", les cas suivants ont été retenus comme éléments rationnels du modèle:

1- La facette d'intérêt ne produit pas de différences marquées dans la difficulté des questions du test; une telle facette ne saurait être qualifiée de catégorie diagnostique.

2- La facette d'intérêt produit à elle-seule des différences marquées dans la difficulté des questions; la difficulté introduite par certains éléments de cette facette est la même sous toutes les conditions créées par d'autres facettes et ce pour la majorité des élèves; il s'agit là d'un premier type de catégorie diagnostique identifié dans le modèle et qui devrait amener un correctif pédagogique de nature collective.

3- La facette d'intérêt dépend de la présence d'autres facettes pour produire des différences en difficulté des questions; dans ce cas, c'est la combinaison de deux ou de plusieurs facettes qui définit une catégorie diagnostique; cet effet combiné de facettes est le même pour la majorité des élèves et, comme dans le cas précédent, le correctif serait de nature collective.

4- La facette d'intérêt produit des différences en difficulté des questions mais pas de la même façon pour tous les élèves; cette facette peut définir une catégorie diagnostique mais dans ce cas son effet ne peut être généralisé à tous les élèves; le correctif à apporter à la non-maîtrise de l'objectif ne saurait alors être de nature collective et il faudrait expliquer ce genre de résultat au moyen de différences individuelles caractérisant le groupe d'élèves choisi comme cible.

Tous les cas possibles n'ont pas été énumérés dans cette présentation du modèle. On comprendra que le modèle devient de plus en plus complexe avec l'augmentation du nombre de facettes.

Le modèle de recherche pédagogique présenté jusqu'ici ne se limite pas à la construction d'un test selon des facettes. D'une part, ce modèle doit être accompagné de l'analyse psychopédagogique de la matière, soit pour anticiper des variations importantes ou pertinentes à introduire dans la construction du test, soit pour expliquer des résultats obtenus. D'autre part, la détection de catégories diagnostiques dans un schème expérimental complexe nous oblige à recourir à des procédés statistiques qui tiennent compte le plus possible de toute la réalité que l'on veut cerner. C'est ainsi que l'analyse des composantes de la variance, puisée à même la théorie de la généralisabilité, est apparue comme étant un outil des plus commodes pour étudier des effets de variations systématiques introduites dans les questions d'une épreuve de rendement.

Les applications de la théorie des facettes et de la théorie de la généralisabilité à la construction d'une épreuve diagnostique de rendement apparaissent inexistantes pour l'instant. Le modèle présenté ici demande donc à être validé au niveau de sa sensibilité et au niveau de sa pertinence. La deuxième partie de cet ouvrage est consacrée aux applications du modèle à des tests déjà construits selon des facettes et pour lesquels des résultats d'analyse de la variance sont accessibles. Cette deuxième partie traitera également de certains points d'ordre méthodologique ou statistique qui mériteraient à être éclaircis pour mieux cerner la portée exacte des conclusions que l'on doit tirer dans la recherche de catégories diagnostiques avec ce modèle.

BIBLIOGRAPHIE

Association des Institutions d'enseignement secondaire (A.I.E.S.), Tests diagnostiques de français: accord du verbe; Montréal: juin 1980,

Airasian, Peter W. et George F. Madaus, "Functional Types of Student Evaluation", Measurement and Evaluation in Guidance, Vol. 4, No 4, 1972, pp. 221-233;

Baker, Eva L., "Beyond Objectives: Domain-Referenced Tests for Evaluation and Instructional Improvement", Educational Technology, juin 1974, pp. 10-16;

Bégin, Yves et Gilles Dussault, "La pédagogie de la maîtrise ou la redécouverte du bon enseignement", Vie pédagogique, Vol. 10, décembre 1980, pp. 4-6;

Bloom, Benjamin S., J. Thomas Hastings et George F. Madaus, Handbook on Formative and Summative Evaluation of Student Learning, New York: McGraw-Hill Book Co., 1971;

Bonboir, Anna, La pédagogie corrective, Paris: Presses Universitaires de France, 1970(a);

-------------, Initiation à l'étude rationnelle du travail des écoliers, Louvain: Vander-éditeur, 1970(b);

Brennan, Robert L. et Michael T. Kane, "Generalizability Theory: A Review", pp. 33-51 dans: Ross Traub (editor), New Directions for Testing and Measurement, San Francisco: Jossey-Bass Inc., Publishers, No 4, 1979;

Campbell, Alison C., "Some Determinants of the Difficulty of Non-Verbal Classification Items", Educational and Psychological Measurement, Vol. XXI, No 4, 1961, pp. 899-913;

-------------------, "Concentration versus Dispersion of Figural Properties in Non-Verbal Test Items", Educational and Psychological Measurement, Vol. XXIV, No 2, 1964, pp. 285-289;

Cardinet, Jean, Yvan Tourneur et Linda Allal, "The Symmetry of Generalizability Theory: Applications to Educational Measurement", Journal of Educational Measurement, Vol. 13, No 2, 1976, pp. 119-135;

Cronbach, Lee J., Goldine C. Gleser, Harinder Nanda et Nageswari Rajaratnam, The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles, New York: John Wiley and Sons, Inc., 1972;

Depover, C., "Contribution à l'étude psychopédagogique du problème d'arithmétique: étude de quelques variables d'énoncé, Scientia Paedagogica Experimentalis, Vol. XVI, 1979, pp. 197-231;

Durnin, John et Joseph M. Scandura, "An Algorithmic Approcah to Assessing Behavior Potential", Journal of Educational Psychology, Vol. 65, 1973, pp. 262-272;

Ebel, Robert L., "Content Standard Test Scores", Educational and Psychological Measurement, Vol. XXII, No 1, 1962, pp. 15-25;

Friend, Jamesine E., Column Addition Skills, communication au congrès annuel de l'American Educational Research Association tenu à San Francisco en avril 1979;

Garon, Suzanne, Laure Ouellet, Aline Savard et Lionel Dion, Analyse du test diagnostique: "Accord du verbe avec son sujet", document miméographié émis par la Commission scolaire régionale du Bas-Saint-Laurent, janvier 1978, 38 pp.;

Glaser, Robert et Anthony J. Nitko, "Measurement in Learning and Instruction", Chapitre 17, pp. 625-670 dans: Robert L. Thorndike, editor, Educational Measurement (2ième édition), Washington D.C., ACE, 1971;

Guttman, Louis, "A New Approach to Factor Analysis: The Radex", dans Lazarsfield, P.F., Mathematical Thinking in the Social Sciences, Glencoe, Illinois: Free Press, 1954;

Guttman, Louis et I.M. Schlesinger, Development of Diagnostic Analytical and Mechanical Ability Tests Through Facet Design and Analysis, Israel Institute of Applied Social Research, Jerisalem, 1966;

----------------------------------, The Analysis of Diagnostic Effectiveness of a Facet Design Battery of Achievement and Analytical Ability Tests, The Israel Institute of Applied Social Research, Jerusalem, 1967;

Haseman, Klauss, "On difficulties with Fractions", Educational Studies in Mathematics, Vol. 12, 1981, pp. 71-87;

Hively, Wells, "Introduction to Domain-Referenced Testing", Educational Technology, Juin 1974, pp. 5-10;

Hively, Wells, Harry L. Patterson et Sara H. Page, "A Universe-defined System of Arithmetic Achievement Tests", Journal of Educational Measurement, Vol. 5, No 4, 1968, pp. 275-290;

Humphreys, Lloyd G., "The Organization of Human Abilities", The American Psychologist, Vol. 17, 1962, pp. 475-483;

Kirk, R.E., Experimental Design: Procedures for the Behavioral Sciences, Belmont, California: Wadsworth, 1968;

Macready, George B. et Jack C. Merwin, "Homogeneity Within Item Forms in Domain-Referenced Testing", Educational and Psychological Measurement, Vol. 33, 1973, pp. 351-360;

Millman, Jason, "Criterion-Referenced Measurement", pp. 311-397 dans: W. James Popham (editor), Evaluation in Education: Current Applications, Berkeley: McCutchan Publishing Co., 1974;

Popham, William J. et T.R. Husek, "Implications of Criterion-Referenced Measurement", pp. 17-37 dans: W. James Popham (editor), Criterion-Referenced Measurement: An Introduction, Englewood Cliffs, New Jersey: Educational Technology Publications, 1971;

Scallon, Gérard, Intrinsic Determinants of the Difficulty of non-verbal Classification Items, thèse de doctorat non publiés, Toronto: Université de Toronto, 1973;

---------------, L'évaluation formative: un outil au service de l'enseignement et de l'apprentissage, document miméographié en dépôt à la Commission scolaire des Mille-Îles, Laval, 1980;

Tourneur, Yvan et Jean Cardinet, Analyse de variance et théorie de la généralisabilité: guide pour la réalisation des calculs, Mons-Neuchatel: document 790.803/CT/9, 1980.

APPENDICE

Carrés moyens attendus et composantes de la variance estimées à partir des carrés moyens observés pour le schème p x a x b dans lequel p (sujets) est aléatoire simple, a et b sont deux facettes du test qui correspondent à des dimensions fixées.

$CMA_{(p)}$: $n_a n_b\ \sigma^2_{(p)}$

$CMA_{(a)}$: $n_b\ \sigma^2_{(pa)} + n_p n_b\ \sigma^2_{(a)}$

$CMA_{(b)}$: $n_a\ \sigma^2_{(pb)} + n_p n_a\ \sigma^2_{(b)}$

$CMA_{(pa)}$: $n_b\ \sigma^2_{(pa)}$

$CMA_{(pb)}$: $n_a\ \sigma^2_{(pb)}$

$CMA_{(ab)}$: $\sigma^2_{(pab)} + n_p\ \sigma^2_{(ab)}$

$CMA_{(pab)}$: $\sigma^2_{(pab)}$

$\hat{\sigma}^2_{(p)}$: $CM_p / n_a n_b$

$\hat{\sigma}^2_{(a)}$: $(CM_a - CM_{pa}) / n_p n_b$ $\hat{\sigma}^2_{(pa)}$: CM_{pa} / n_b

$\hat{\sigma}^2_{(b)}$: $(CM_b - CM_{pb}) / n_p n_a$ $\hat{\sigma}^2_{(pb)}$: CM_{pb} / n_a

$\hat{\sigma}^2_{(ab)}$: $(CM_{ab} - CM_{pab}) / n_p$

$\hat{\sigma}^2_{(pab)}$: CM_{pab}

Responsable des éditions: Marc-André Nadeau Ph.D.
Co-responsable: Richard Girard Ph.D.

Rédacteurs:

Jean-Jacques Bernier Ph.D. Université Laval
Vidya Bhushan Ph.D. Université Laval
François Dupuis Ph.D. Université Laval
Richard Girard Ph.D. Université Laval
Marc-André Nadeau Ph.D. Université Laval
Gérard Scallon Ph.D. Université Laval
Claude Valiquette M.A. Université Laval
Jean-Paul Voyer Ph.D. Université Laval

Comité de lecture:

Denis Allaire Université de Sherbrooke
Roger Asselin U.Q.T.R. Trois-Rivières
Yves Bégin I.N.R.S. Éducation
Richard Bertrand I.N.R.S. Éducation
Marven W. Boss Université d'Ottawa
Fernand Boucher Université de Montréal
André Caillé U.Q.A.C. Chicoutimi
Robert Côté Université Laval
Clément Dassa Université de Montréal
Jean Dulude S.P.U. Université de Montréal
Jacques Émond Ministère de l'éducation
Marie-Chantal Guédon Université de Montréal
Guy Huot Montréal
Hubert Laforge Université Laval
Louis Laurencelle U.Q.T.R. Trois-Rivières
Guy Lavoie Université de Montréal
Martin Morf Université de Windsor
Gilles G. Nadeau Université de Moncton
Jacques Perron Université de Montréal
Jacques Plante S.P.U. Université Laval
Serge Racine Université de Montréal
Romain Rousseau U.Q.A.R. Rimouski
Serge Séguin U.Q.A.M. Montréal
Bernard Tétreau Université de Montréal
Robert Tousignant Université de Montréal
Michel Trahan Université de Montréal
Suzanne Winsberg Université de Montréal